GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: January 2, 2003, 15:06:47 ; Search time 3759 Seconds
(without alignments)
2688.822 Million cell updates/sec

Title: US-09-850-147-1
Perfect score: 402
Sequence: 1 ccgggccgggctcaccagct.....ggcataacttcctgtgttcc 402

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 24791104 seqs, 12571243825 residues

Word size : 0

Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/1/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/1/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/1/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/1/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/1/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/1/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/1/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/1/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/1/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/1/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/1/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/1/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/1/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/1/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/1/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/1/pna/US092_COMB.seq:*
17: /cgn2_6/ptodata/1/pna/US093_COMB.seq:*
18: /cgn2_6/ptodata/1/pna/US094_COMB.seq:*
19: /cgn2_6/ptodata/1/pna/US095A_COMB.seq:*
20: /cgn2_6/ptodata/1/pna/US095B_COMB.seq:*
21: /cgn2_6/ptodata/1/pna/US095C_COMB.seq:*
22: /cgn2_6/ptodata/1/pna/US095D_COMB.seq:*
23: /cgn2_6/ptodata/1/pna/US096A_COMB.seq:*
24: /cgn2_6/ptodata/1/pna/US096B_COMB.seq:*
25: /cgn2_6/ptodata/1/pna/US096C_COMB.seq:*
26: /cgn2_6/ptodata/1/pna/US096D_COMB.seq:*
27: /cgn2_6/ptodata/1/pna/US096E_COMB.seq:*
28: /cgn2_6/ptodata/1/pna/US097A_COMB.seq:*
29: /cgn2_6/ptodata/1/pna/US097B_COMB.seq:*
30: /cgn2_6/ptodata/1/pna/US097C_COMB.seq:*
31: /cgn2_6/ptodata/1/pna/US098A_COMB.seq:*
32: /cgn2_6/ptodata/1/pna/US098B_COMB.seq:*
33: /cgn2_6/ptodata/1/pna/US098C_COMB.seq:*
34: /cgn2_6/ptodata/1/pna/US099A_COMB.seq:*
35: /cgn2_6/ptodata/1/pna/US099B_COMB.seq:*
36: /cgn2_6/ptodata/1/pna/US099C_COMB.seq:*
37: /cgn2_6/ptodata/1/pna/US099D_COMB.seq:*
38: /cgn2_6/ptodata/1/pna/US100A_COMB.seq:*
39: /cgn2_6/ptodata/1/pna/US100B_COMB.seq:*
40: /cgn2_6/ptodata/1/pna/US101A_COMB.seq:*
41: /cgn2_6/ptodata/1/pna/US101B_COMB.seq:*
42: /cgn2_6/ptodata/1/pna/US102A_COMB.seq:*
43: /cgn2_6/ptodata/1/pna/US102B_COMB.seq:*
44: /cgn2_6/ptodata/1/pna/US6000_COMB.seq:*
45: /cgn2_6/ptodata/1/pna/US6001_COMB.seq:*
46: /cgn2_6/ptodata/1/pna/US6002_COMB.seq:*
47: /cgn2_6/ptodata/1/pna/US6003_COMB.seq:*
48: /cgn2_6/ptodata/1/pna/US6004_COMB.seq:*
49: /cgn2_6/ptodata/1/pna/US6005_COMB.seq:*
50: /cgn2_6/ptodata/1/pna/US6006_COMB.seq:*
51: /cgn2_6/ptodata/1/pna/US6007_COMB.seq:*
52: /cgn2_6/ptodata/1/pna/US6008_COMB.seq:*
53: /cgn2_6/ptodata/1/pna/US6009_COMB.seq:*
54: /cgn2_6/ptodata/1/pna/US6010_COMB.seq:*
55: /cgn2_6/ptodata/1/pna/US6011_COMB.seq:*
56: /cgn2_6/ptodata/1/pna/US6012_COMB.seq:*
57: /cgn2_6/ptodata/1/pna/US6013_COMB.seq:*
58: /cgn2_6/ptodata/1/pna/US6014_COMB.seq:*
59: /cgn2_6/ptodata/1/pna/US6015_COMB.seq:*
60: /cgn2_6/ptodata/1/pna/US6016_COMB.seq:*
61: /cgn2_6/ptodata/1/pna/US6017_COMB.seq:*
62: /cgn2_6/ptodata/1/pna/US6018_COMB.seq:*
63: /cgn2_6/ptodata/1/pna/US6019_COMB.seq:*
64: /cgn2_6/ptodata/1/pna/US6020_COMB.seq:*
65: /cgn2_6/ptodata/1/pna/US6021_COMB.seq:*
66: /cgn2_6/ptodata/1/pna/US6022_COMB.seq:*
67: /cgn2_6/ptodata/1/pna/US6023_COMB.seq:*
68: /cgn2_6/ptodata/1/pna/US6024_COMB.seq:*
69: /cgn2_6/ptodata/1/pna/US6025_COMB.seq:*
70: /cgn2_6/ptodata/1/pna/US6026_COMB.seq:*
71: /cgn2_6/ptodata/1/pna/US6027_COMB.seq:*
72: /cgn2_6/ptodata/1/pna/US6028_COMB.seq:*
73: /cgn2_6/ptodata/1/pna/US6029_COMB.seq:*
74: /cgn2_6/ptodata/1/pna/US6030_COMB.seq:*
75: /cgn2_6/ptodata/1/pna/US6031_COMB.seq:*
76: /cgn2_6/ptodata/1/pna/US6032_COMB.seq:*
77: /cgn2_6/ptodata/1/pna/US6033_COMB.seq:*
78: /cgn2_6/ptodata/1/pna/US6034_COMB.seq:*
79: /cgn2_6/ptodata/1/pna/US6035_COMB.seq:*
80: /cgn2_6/ptodata/1/pna/US6036_COMB.seq:*
81: /cgn2_6/ptodata/1/pna/US6037_COMB.seq:*
82: /cgn2_6/ptodata/1/pna/US6038_COMB.seq:*
83: /cgn2_6/ptodata/1/pna/US6039_COMB.seq:*
84: /cgn2_6/ptodata/1/pna/US6040_COMB.seq:*
85: /cgn2_6/ptodata/1/pna/US6041_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	402	100.0	402	25	US-09-654-617-457554	Sequence 457554,
	2	402	100.0	402	27	US-09-684-016-457554	Sequence 457554,
	3	402	100.0	402	32	US-09-850-147-1	Sequence 1, Appli
	4	402	100.0	402	64	US-60-202-213-1	Sequence 1, Appli
c	5	29	7.2	341	33	US-09-865-439A-8889	Sequence 8889, Ap
c	6	29	7.2	351	33	US-09-865-439A-4361	Sequence 4361, Ap
	7	29	7.2	385	32	US-09-850-147-10	Sequence 10, Appl
	8	29	7.2	385	64	US-60-202-213-10	Sequence 10, Appl
c	9	29	7.2	388	33	US-09-865-439A-18142	Sequence 18142, A
	10	29	7.2	393	22	US-09-565-306-82924	Sequence 82924, A
	11	29	7.2	426	32	US-09-850-147-40	Sequence 40, Appl
	12	29	7.2	426	64	US-60-202-213-40	Sequence 40, Appl
	13	29	7.2	433	32	US-09-850-147-1092	Sequence 1092, Ap
	14	29	7.2	433	64	US-60-202-213-1089	Sequence 1089, Ap
c	15	29	7.2	450	64	US-60-207-458-54563	Sequence 54563, A
	16	29	7.2	456	25	US-09-654-617-451847	Sequence 451847,
	17	29	7.2	456	27	US-09-684-016-451847	Sequence 451847,
c	18	29	7.2	484	33	US-09-865-419A-15731	Sequence 15731, A
c	19	29	7.2	484	64	US-60-208-063-2535	Sequence 2535, Ap
c	20	29	7.2	545	33	US-09-873-402A-41873	Sequence 41873, A
c	21	29	7.2	550	33	US-09-865-419A-51912	Sequence 51912, A

c	22	29	7.2	559	33	US-09-873-402A-54657	Sequence 54657, A
c	23	29	7.2	559	64	US-60-209-830-46016	Sequence 46016, A
c	24	29	7.2	570	33	US-09-865-439A-116358	Sequence 116358,
	25	29	7.2	605	33	US-09-865-439A-93607	Sequence 93607, A
	26	29	7.2	605	64	US-60-207-458-137622	Sequence 137622,
	27	29	7.2	660	33	US-09-861-478-52	Sequence 52, Appl
c	28	29	7.2	675	33	US-09-861-478-506	Sequence 506, App
	29	29	7.2	995	42	US-10-219-999-1500	Sequence 1500, Ap
	30	29	7.2	1231	42	US-10-219-999-20024	Sequence 20024, A
	31	29	7.2	1325	42	US-10-219-999-18314	Sequence 18314, A
	32	28	7.0	1694	25	US-09-654-617-262628	Sequence 262628,
	33	28	7.0	1694	27	US-09-684-016-262628	Sequence 262628,
c	34	27	6.7	345	64	US-60-207-458-47364	Sequence 47364, A
c	35	24	6.0	397	64	US-60-207-458-43952	Sequence 43952, A
	36	22	5.5	262	16	US-09-244-000A-14941	Sequence 14941, A
	37	22	5.5	262	36	US-09-978-703-14941	Sequence 14941, A
c	38	22	5.5	349	19	US-09-521-640-256139	Sequence 256139,
c	39	22	5.5	349	58	US-60-140-769-43520	Sequence 43520, A
c	40	22	5.5	396	19	US-09-521-640-144941	Sequence 144941,
c	41	22	5.5	396	19	US-09-521-640-156054	Sequence 156054,
c	42	22	5.5	411	33	US-09-865-419A-9944	Sequence 9944, Ap
	43	22	5.5	574	24	US-09-630-346-335	Sequence 335, App
	44	22	5.5	574	39	US-10-062-254-335	Sequence 335, App
	45	22	5.5	607	33	US-09-874-708A-77966	Sequence 77966, A

ALIGNMENTS

RESULT 1
US-09-654-617-457554
; Sequence 457554, Application US/09654617
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; TITLE OF INVENTION: Annotated Plant Genes
; FILE REFERENCE: 38-21(15097)D
; CURRENT APPLICATION NUMBER: US/09/654,617
; CURRENT FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 463173
; SEQ ID NO 457554
; LENGTH: 402
; TYPE: DNA
; ORGANISM: Sorghum bicolor
US-09-654-617-457554

Query Match 100.0%; Score 402; DB 25; Length 402;
Best Local Similarity 100.0%; Pred. No. 2.9e-212;
Matches 402; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 CCGGGCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CCGGGCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGG 60

Qy     61 CCTTGAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCTTGAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTAT 120

Qy    121 ATTAAGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ATTAAGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAA 180

Qy    181 ACAACCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGAC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ACAACCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGAC 240

Qy    241 AATTTTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AATTTTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATT 300

Qy    301 GTGGAGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GTGGAGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACA 360

Qy    361 AGTACAAAAAACCAACGGGAGGGGCATAACTTCCTGTGTTCC 402
          |||||||||||||||||||||||||||||||||||||||||||
Db    361 AGTACAAAAAACCAACGGGAGGGGCATAACTTCCTGTGTTCC 402
```

RESULT 2
US-09-684-016-457554
; Sequence 457554, Application US/09684016
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; TITLE OF INVENTION: Annotated Plant Genes
; FILE REFERENCE: 38-21(15097)D
; CURRENT APPLICATION NUMBER: US/09/684,016
; CURRENT FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: US 09/654,617
; PRIOR FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 463173
; SEQ ID NO 457554
; LENGTH: 402
; TYPE: DNA
; ORGANISM: Sorghum bicolor
US-09-684-016-457554

Query Match 100.0%; Score 402; DB 27; Length 402;
Best Local Similarity 100.0%; Pred. No. 2.9e-212;
Matches 402; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 CCGGGCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CCGGGCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGG 60

Qy     61 CCTTGAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCTTGAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTAT 120

Qy    121 ATTAAGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ATTAAGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAA 180

Qy    181 ACAACCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGAC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ACAACCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGAC 240

Qy    241 AATTTTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AATTTTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATT 300

Qy    301 GTGGAGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GTGGAGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACA 360

Qy    361 AGTACAAAAAACCAACGGGAGGGGCATAACTTCCTGTGTTCC 402
          |||||||||||||||||||||||||||||||||||||||||||
Db    361 AGTACAAAAAACCAACGGGAGGGGCATAACTTCCTGTGTTCC 402
```

RESULT 3
US-09-850-147-1
; Sequence 1, Application US/09850147
; GENERAL INFORMATION:
; APPLICANT: Andersen, Scott E.
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(51914)B
; CURRENT APPLICATION NUMBER: US/09/850,147
; CURRENT FILING DATE: 2001-05-08
; PRIOR APPLICATION NUMBER: US 60/202,213
; PRIOR FILING DATE: 2000-05-08
; PRIOR APPLICATION NUMBER: US 09/654,617
; PRIOR FILING DATE: 2000-09-05

```
;  PRIOR APPLICATION NUMBER: US 09/684,016
;  PRIOR FILING DATE: 2000-10-10
;  NUMBER OF SEQ ID NOS: 18014
; SEQ ID NO 1
;   LENGTH: 402
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   OTHER INFORMATION: Clone ID: LIB3476-001-Q6-K1-A3
US-09-850-147-1

  Query Match             100.0%;  Score 402;  DB 32;  Length 402;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-212;
  Matches  402;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CCGGGCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CCGGGCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGG 60

Qy     61 CCTTGAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCTTGAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTAT 120

Qy    121 ATTAAGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ATTAAGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAA 180

Qy    181 ACAACCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGAC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ACAACCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGAC 240

Qy    241 AATTTTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AATTTTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATT 300

Qy    301 GTGGAGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GTGGAGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACA 360

Qy    361 AGTACAAAAAACCAACGGGAGGGGCATAACTTCCTGTGTTCC 402
          |||||||||||||||||||||||||||||||||||||||||||
Db    361 AGTACAAAAAACCAACGGGAGGGGCATAACTTCCTGTGTTCC 402


RESULT 4
US-60-202-213-1
; Sequence 1, Application US/60202213
; GENERAL INFORMATION:
;  APPLICANT:       Andersen, Scott E.
;  APPLICANT:  Edgerton, Michael D
;  TITLE OF INVENTION:       NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
;  TITLE OF INVENTION:  PLANTS
;  FILE REFERENCE:       38-21(51914)A
;  CURRENT APPLICATION NUMBER: US/60/202,213
;  CURRENT FILING DATE:  2000-05-08
;  NUMBER OF SEQ ID NOS:       17986
; SEQ ID NO 1
;   LENGTH: 402
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   OTHER INFORMATION: Clone ID: LIB3476-001-Q6-K1-A3
US-60-202-213-1

  Query Match             100.0%;  Score 402;  DB 64;  Length 402;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-212;
  Matches  402;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CCGGGCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CCGGGCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGG 60

Qy     61 CCTTGAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCTTGAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTAT 120

Qy    121 ATTAAGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ATTAAGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAA 180

Qy    181 ACAACCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGAC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ACAACCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGAC 240

Qy    241 AATTTTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AATTTTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATT 300

Qy    301 GTGGAGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GTGGAGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACA 360

Qy    361 AGTACAAAAAACCAACGGGAGGGGCATAACTTCCTGTGTTCC 402
          |||||||||||||||||||||||||||||||||||||||||||
Db    361 AGTACAAAAAACCAACGGGAGGGGCATAACTTCCTGTGTTCC 402


RESULT 5
US-09-865-439A-8889/c
; Sequence 8889, Application US/09865439A
; GENERAL INFORMATION:
;  APPLICANT: Edgerton, Michael D
;  APPLICANT:  Hardeman, Kristine J.
;  APPLICANT:  La Rosa, Thomas J.
;  TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(51936)B
;  CURRENT APPLICATION NUMBER: US/09/865,439A
;  CURRENT FILING DATE:  2001-05-29
;  PRIOR APPLICATION NUMBER: US 60/207,458
;  PRIOR FILING DATE: 2000-05-30
;  NUMBER OF SEQ ID NOS: 119126
; SEQ ID NO 8889
;   LENGTH: 341
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3205-106-P1-N1-B3
US-09-865-439A-8889

  Query Match              7.2%;  Score 29;  DB 33;  Length 341;
  Best Local Similarity   100.0%;  Pred. No. 0.00015;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    112 AGCTTCTATATTAAGCTTCGATGAGAAGA 140
          ||||||||||||||||||||||||||||||
Db    336 AGCTTCTATATTAAGCTTCGATGAGAAGA 308


RESULT 6
US-09-865-439A-4361/c
; Sequence 4361, Application US/09865439A
; GENERAL INFORMATION:
;  APPLICANT: Edgerton, Michael D
;  APPLICANT:  Hardeman, Kristine J.
;  APPLICANT:  La Rosa, Thomas J.
;  TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(51936)B
;  CURRENT APPLICATION NUMBER: US/09/865,439A
;  CURRENT FILING DATE:  2001-05-29
;  PRIOR APPLICATION NUMBER: US 60/207,458
;  PRIOR FILING DATE: 2000-05-30
;  NUMBER OF SEQ ID NOS: 119126
; SEQ ID NO 4361
;   LENGTH: 351
;   TYPE: DNA
;   ORGANISM: Zea mays
```

```
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3205-049-P1-N1-F9
US-09-865-439A-4361

  Query Match             7.2%;  Score 29;  DB 33;  Length 351;
  Best Local Similarity   100.0%;  Pred. No. 0.00015;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    112 AGCTTCTATATTAAGCTTCGATGAGAAGA 140
          ||||||||||||||||||||||||||||||
Db    277 AGCTTCTATATTAAGCTTCGATGAGAAGA 249


RESULT 7
US-09-850-147-10
; Sequence 10, Application US/09850147
; GENERAL INFORMATION:
;  APPLICANT: Andersen, Scott E.
;  APPLICANT:  Edgerton, Michael D
;  TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(51914)B
;  CURRENT APPLICATION NUMBER: US/09/850,147
;  CURRENT FILING DATE:  2001-05-08
;  PRIOR APPLICATION NUMBER: US 60/202,213
;  PRIOR FILING DATE: 2000-05-08
;  PRIOR APPLICATION NUMBER: US 09/654,617
;  PRIOR FILING DATE: 2000-09-05
;  PRIOR APPLICATION NUMBER: US 09/684,016
;  PRIOR FILING DATE: 2000-10-10
;  NUMBER OF SEQ ID NOS: 18014
; SEQ ID NO 10
;   LENGTH: 385
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   OTHER INFORMATION: Clone ID: LIB3476-001-Q6-K2-A3
US-09-850-147-10

  Query Match             7.2%;  Score 29;  DB 32;  Length 385;
  Best Local Similarity   100.0%;  Pred. No. 0.00015;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    112 AGCTTCTATATTAAGCTTCGATGAGAAGA 140
          ||||||||||||||||||||||||||||||
Db     96 AGCTTCTATATTAAGCTTCGATGAGAAGA 124


RESULT 8
US-60-202-213-10
; Sequence 10, Application US/60202213
; GENERAL INFORMATION:
;  APPLICANT:       Andersen, Scott E.
;  APPLICANT:  Edgerton, Michael D
;  TITLE OF INVENTION:       NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
;  TITLE OF INVENTION:  PLANTS
;  FILE REFERENCE:       38-21(51914)A
;  CURRENT APPLICATION NUMBER: US/60/202,213
;  CURRENT FILING DATE:  2000-05-08
;  NUMBER OF SEQ ID NOS:       17986
; SEQ ID NO 10
;   LENGTH: 385
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   OTHER INFORMATION: Clone ID: LIB3476-001-Q6-K2-A3
US-60-202-213-10

  Query Match             7.2%;  Score 29;  DB 64;  Length 385;
  Best Local Similarity   100.0%;  Pred. No. 0.00015;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    112 AGCTTCTATATTAAGCTTCGATGAGAAGA 140
          ||||||||||||||||||||||||||||||
Db     96 AGCTTCTATATTAAGCTTCGATGAGAAGA 124


RESULT 9
US-09-865-439A-18142/c
; Sequence 18142, Application US/09865439A
; GENERAL INFORMATION:
;  APPLICANT: Edgerton, Michael D
;  APPLICANT:  Hardeman, Kristine J.
;  APPLICANT:  La Rosa, Thomas J.
;  TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(51936)B
;  CURRENT APPLICATION NUMBER: US/09/865,439A
;  CURRENT FILING DATE:  2001-05-29
;  PRIOR APPLICATION NUMBER: US 60/207,458
;  PRIOR FILING DATE: 2000-05-30
;  NUMBER OF SEQ ID NOS: 119126
; SEQ ID NO 18142
;   LENGTH: 388
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3206-089-P1-N1-H4
US-09-865-439A-18142

  Query Match             7.2%;  Score 29;  DB 33;  Length 388;
  Best Local Similarity   100.0%;  Pred. No. 0.00015;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    112 AGCTTCTATATTAAGCTTCGATGAGAAGA 140
          ||||||||||||||||||||||||||||||
Db    333 AGCTTCTATATTAAGCTTCGATGAGAAGA 305


RESULT 10
US-09-565-306-82924
; Sequence 82924, Application US/09565306
; GENERAL INFORMATION:
;  APPLICANT:       Andersen, Scott E.
;  APPLICANT:  Conner, Timothy W.
;  APPLICANT:  Lalgudi, Raghunath V.
;  TITLE OF INVENTION:       Nucleic Acid Molecules and Other Molecules Associated wi
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE:       38-21(15459)C
;  CURRENT APPLICATION NUMBER: US/09/565,306
;  CURRENT FILING DATE:  2000-05-04
;  NUMBER OF SEQ ID NOS:       83523
; SEQ ID NO 82924
;   LENGTH: 393
;   TYPE: DNA
;   ORGANISM: Zea mays
;   OTHER INFORMATION: Clone ID: LIB3114-050-P1-K1-G7
US-09-565-306-82924

  Query Match             7.2%;  Score 29;  DB 22;  Length 393;
  Best Local Similarity   100.0%;  Pred. No. 0.00015;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    112 AGCTTCTATATTAAGCTTCGATGAGAAGA 140
          ||||||||||||||||||||||||||||||
Db    110 AGCTTCTATATTAAGCTTCGATGAGAAGA 138


RESULT 11
US-09-850-147-40
; Sequence 40, Application US/09850147
; GENERAL INFORMATION:
;  APPLICANT: Andersen, Scott E.
;  APPLICANT:  Edgerton, Michael D
;  TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(51914)B
;  CURRENT APPLICATION NUMBER: US/09/850,147
```

```
;  CURRENT FILING DATE:  2001-05-08
;  PRIOR APPLICATION NUMBER: US 60/202,213
;  PRIOR FILING DATE: 2000-05-08
;  PRIOR APPLICATION NUMBER: US 09/654,617
;  PRIOR FILING DATE: 2000-09-05
;  PRIOR APPLICATION NUMBER: US 09/684,016
;  PRIOR FILING DATE: 2000-10-10
;  NUMBER OF SEQ ID NOS: 18014
; SEQ ID NO 40
;   LENGTH: 426
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   OTHER INFORMATION: Clone ID: LIB3476-004-Q6-K1-A3
US-09-850-147-40

  Query Match             7.2%;  Score 29;  DB 32;  Length 426;
  Best Local Similarity   100.0%;  Pred. No. 0.00015;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    112 AGCTTCTATATTAAGCTTCGATGAGAAGA 140
          ||||||||||||||||||||||||||||||
Db    135 AGCTTCTATATTAAGCTTCGATGAGAAGA 163


RESULT 12
US-60-202-213-40
; Sequence 40, Application US/60202213
; GENERAL INFORMATION:
;  APPLICANT:       Andersen, Scott E.
;  APPLICANT:  Edgerton, Michael D
;  TITLE OF INVENTION:       NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
;  TITLE OF INVENTION:  PLANTS
;  FILE REFERENCE:       38-21(51914)A
;  CURRENT APPLICATION NUMBER: US/60/202,213
;  CURRENT FILING DATE:  2000-05-08
;  NUMBER OF SEQ ID NOS:       17986
; SEQ ID NO 40
;   LENGTH: 426
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   OTHER INFORMATION: Clone ID: LIB3476-004-Q6-K1-A3
US-60-202-213-40

  Query Match             7.2%;  Score 29;  DB 64;  Length 426;
  Best Local Similarity   100.0%;  Pred. No. 0.00015;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    112 AGCTTCTATATTAAGCTTCGATGAGAAGA 140
          ||||||||||||||||||||||||||||||
Db    135 AGCTTCTATATTAAGCTTCGATGAGAAGA 163


RESULT 13
US-09-850-147-1092
; Sequence 1092, Application US/09850147
; GENERAL INFORMATION:
;  APPLICANT: Andersen, Scott E.
;  APPLICANT:  Edgerton, Michael D
;  TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(51914)B
;  CURRENT APPLICATION NUMBER: US/09/850,147
;  CURRENT FILING DATE:  2001-05-08
;  PRIOR APPLICATION NUMBER: US 60/202,213
;  PRIOR FILING DATE: 2000-05-08
;  PRIOR APPLICATION NUMBER: US 09/654,617
;  PRIOR FILING DATE: 2000-09-05
;  PRIOR APPLICATION NUMBER: US 09/684,016
;  PRIOR FILING DATE: 2000-10-10
;  NUMBER OF SEQ ID NOS: 18014
; SEQ ID NO 1092
;   LENGTH: 433
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(433)
;   OTHER INFORMATION: unsure at all n locations
;   OTHER INFORMATION: Clone ID: LIB3476-004-Q6-K2-A3
US-09-850-147-1092

  Query Match             7.2%;  Score 29;  DB 32;  Length 433;
  Best Local Similarity   100.0%;  Pred. No. 0.00015;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    112 AGCTTCTATATTAAGCTTCGATGAGAAGA 140
          ||||||||||||||||||||||||||||||
Db    105 AGCTTCTATATTAAGCTTCGATGAGAAGA 133


RESULT 14
US-60-202-213-1089
; Sequence 1089, Application US/60202213
; GENERAL INFORMATION:
;  APPLICANT:       Andersen, Scott E.
;  APPLICANT:  Edgerton, Michael D
;  TITLE OF INVENTION:       NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WI
;  TITLE OF INVENTION:  PLANTS
;  FILE REFERENCE:       38-21(51914)A
;  CURRENT APPLICATION NUMBER: US/60/202,213
;  CURRENT FILING DATE:  2000-05-08
;  NUMBER OF SEQ ID NOS:       17986
; SEQ ID NO 1089
;   LENGTH: 433
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   OTHER INFORMATION: Clone ID: LIB3476-004-Q6-K2-A3
US-60-202-213-1089

  Query Match             7.2%;  Score 29;  DB 64;  Length 433;
  Best Local Similarity   100.0%;  Pred. No. 0.00015;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    112 AGCTTCTATATTAAGCTTCGATGAGAAGA 140
          ||||||||||||||||||||||||||||||
Db    105 AGCTTCTATATTAAGCTTCGATGAGAAGA 133


RESULT 15
US-60-207-458-54563/c
; Sequence 54563, Application US/60207458
; GENERAL INFORMATION:
;  APPLICANT:       Abad, Mark S.
;  APPLICANT:  Conner, Timothy W.
;  APPLICANT:  Deikman, Jill
;  APPLICANT:  Hardeman, Kristine J.
;  APPLICANT:  La Rosa, Thomas J.
;  APPLICANT:  Lalgudi, Raghunath V.
;  APPLICANT:  Ruan, Yijun G.
;  APPLICANT:  Ruff, Thomas G.
;  APPLICANT:  Sammons, R. Douglas
;  APPLICANT:  Shukla, Hridayabhiranjan
;  APPLICANT:  Wu, Kunsheng
;  APPLICANT:  Xu, Nanfei
;  TITLE OF INVENTION:       NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WI
;  TITLE OF INVENTION:  PLANTS
;  FILE REFERENCE:       38-21(51936)A
;  CURRENT APPLICATION NUMBER: US/60/207,458
;  CURRENT FILING DATE:  2000-05-30
;  NUMBER OF SEQ ID NOS:       152403
; SEQ ID NO 54563
;   LENGTH: 450
;   TYPE: DNA
;   ORGANISM: Zea mays
;   OTHER INFORMATION: Clone ID: uC-zmflb731234e12a1
US-60-207-458-54563
```

Query Match 7.2%; Score 29; DB 64; Length 450;
Best Local Similarity 100.0%; Pred. No. 0.00015;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy     112 AGCTTCTATATTAAGCTTCGATGAGAAGA 140
           ||||||||||||||||||||||||||||||
Db     321 AGCTTCTATATTAAGCTTCGATGAGAAGA 293
```

Search completed: January 2, 2003, 17:47:03
Job time : 3761 secs

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: January 2, 2003, 13:24:21 ; Search time 3200 Seconds
(without alignments)
3656.040 Million cell updates/sec

Title: US-09-850-147-1
Perfect score: 402
Sequence: 1 ccgggccgggctcaccagct.....ggcataacttcctgtgttcc 402

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		23	5.7	173852	2	AC095235	AC095235 Rattus no
2		23	5.7	176340	2	AC120750	AC120750 Rattus no
3	c	20	5.0	369	9	HUMCOL14	U76577 Human 180 k
4	c	20	5.0	2964	1	AF063186	AF063186 Arthrobac
5		20	5.0	37000	8	SPAC19G12	Z97209 S.pombe chr
6	c	20	5.0	59237	2	AC100210	AC100210 Mus muscu
7	c	20	5.0	60748	2	AC119163	AC119163 Mus muscu
8	c	20	5.0	84767	9	AC004149	AC004149 Homo sapi
9	c	20	5.0	128459	2	AC091230	AC091230 Homo sapi
10		20	5.0	141970	9	AL138761	AL138761 Human DNA
11	c	20	5.0	143620	9	AP001883	AP001883 Homo sapi
12	c	20	5.0	145695	2	AC015690	AC015690 Homo sapi
13		20	5.0	145870	2	AC068675	AC068675 Homo sapi
14	c	20	5.0	151937	9	AP001980	AP001980 Homo sapi
15		20	5.0	156186	2	AC036150	AC036150 Homo sapi
16	c	20	5.0	165315	2	AC109304	AC109304 Mus muscu
17	c	20	5.0	177744	2	AC073954	AC073954 Homo sapi
18	c	20	5.0	198446	2	AC123798	AC123798 Mus muscu
19	c	20	5.0	208979	2	AC023610	AC023610 Mus muscu
20	c	20	5.0	221576	2	AL731822	AL731822 Mus muscu
21		20	5.0	230160	2	AC020705	AC020705 Homo sapi
22		20	5.0	241585	2	AL844536	AL844536 Mus muscu
23		19	4.7	324	6	AX246754	AX246754 Sequence
24		19	4.7	615	6	AX389622	AX389622 Sequence
25		19	4.7	3519	8	ATH297264	AJ297264 Arabidops
26		19	4.7	3618	8	AF412407	AF412407 Arabidops
27		19	4.7	3786	8	AY096796	AY096796 Arabidops
28	c	19	4.7	23273	9	AC093389	AC093389 Homo sapi
29		19	4.7	45692	9	AC108153	AC108153 Homo sapi
30	c	19	4.7	49338	2	AC101240	AC101240 Mus muscu
31	c	19	4.7	49393	2	AC101083	AC101083 Mus muscu
32		19	4.7	53707	9	AC116359	AC116359 Homo sapi
33	c	19	4.7	54049	9	AC110089	AC110089 Homo sapi
34		19	4.7	56750	2	AC106014	AC106014 Homo sapi
35	c	19	4.7	57671	2	AC100231	AC100231 Mus muscu
36		19	4.7	61524	9	AL353725	AL353725 Human DNA
37	c	19	4.7	69948	9	AC016674	AC016674 Homo sapi
38		19	4.7	75412	2	AC015489	AC015489 Homo sapi
39		19	4.7	81662	8	AB008265	AB008265 Arabidops
40		19	4.7	84952	2	AC015487	AC015487 Homo sapi
41		19	4.7	94414	9	AC051663	AC051663 Homo sapi
42		19	4.7	98339	2	AC114128	AC114128 Rattus no
43		19	4.7	105528	8	AC002392	AC002392 Arabidops
44		19	4.7	109321	9	AC021079	AC021079 Homo sapi
45		19	4.7	114491	9	AC010331	AC010331 Homo sapi

ALIGNMENTS

RESULT 1
AC095235
LOCUS AC095235 173852 bp DNA linear HTG 11-JUL-2002
DEFINITION Rattus norvegicus clone CH230-10F21, *** SEQUENCING IN PROGRESS ***, 79 unordered pieces.
ACCESSION AC095235
VERSION AC095235.3 GI:21722866
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 173852)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-ósman,F.R., Allen,C.,

```
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 173852)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 173852)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 10, 2002 this sequence version replaced gi:17942490.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GDMA
                Center clone name: CH230-10F21
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 92252 bases at least Q40
                Consensus quality: 103231 bases at least Q30
                Consensus quality: 111005 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 79 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1160: contig of 1160 bp in length
            *    1161     1260: gap of unknown length
            *    1261     2358: contig of 1098 bp in length
            *    2359     2458: gap of unknown length
            *    2459     3591: contig of 1133 bp in length
            *    3592     3691: gap of unknown length
            *    3692     4862: contig of 1171 bp in length
            *    4863     4962: gap of unknown length
            *    4963     6091: contig of 1129 bp in length
            *    6092     6191: gap of unknown length
            *    6192     7215: contig of 1024 bp in length
            *    7216     7315: gap of unknown length
            *    7316     8481: contig of 1166 bp in length
            *    8482     8581: gap of unknown length
            *    8582     9653: contig of 1072 bp in length
            *    9654     9753: gap of unknown length
            *    9754    11009: contig of 1256 bp in length
            *   11010    11109: gap of unknown length
            *   11110    12257: contig of 1148 bp in length
            *   12258    12357: gap of unknown length
            *   12358    13387: contig of 1030 bp in length
            *   13388    13487: gap of unknown length
            *   13488    14672: contig of 1185 bp in length
            *   14673    14772: gap of unknown length
            *   14773    15803: contig of 1031 bp in length
            *   15804    15903: gap of unknown length
            *   15904    17247: contig of 1344 bp in length
            *   17248    17347: gap of unknown length
            *   17348    18691: contig of 1344 bp in length
            *   18692    18791: gap of unknown length
            *   18792    19920: contig of 1129 bp in length
            *   19921    20020: gap of unknown length
            *   20021    21347: contig of 1327 bp in length
            *   21348    21447: gap of unknown length
            *   21448    22532: contig of 1085 bp in length
            *   22533    22632: gap of unknown length
            *   22633    23647: contig of 1015 bp in length
            *   23648    23747: gap of unknown length
            *   23748    25062: contig of 1315 bp in length
            *   25063    25162: gap of unknown length
            *   25163    26323: contig of 1161 bp in length
            *   26324    26423: gap of unknown length
            *   26424    27991: contig of 1568 bp in length
            *   27992    28091: gap of unknown length
            *   28092    29133: contig of 1042 bp in length
            *   29134    29233: gap of unknown length
            *   29234    30658: contig of 1425 bp in length
            *   30659    30758: gap of unknown length
            *   30759    31952: contig of 1194 bp in length
            *   31953    32052: gap of unknown length
            *   32053    33165: contig of 1113 bp in length
            *   33166    33265: gap of unknown length
            *   33266    34716: contig of 1451 bp in length
            *   34717    34816: gap of unknown length
            *   34817    36686: contig of 1870 bp in length
            *   36687    36786: gap of unknown length
            *   36787    38025: contig of 1239 bp in length
            *   38026    38125: gap of unknown length
            *   38126    39802: contig of 1677 bp in length
            *   39803    39902: gap of unknown length
            *   39903    42910: contig of 3008 bp in length
            *   42911    43010: gap of unknown length
            *   43011    44211: contig of 1201 bp in length
            *   44212    44311: gap of unknown length
            *   44312    45472: contig of 1161 bp in length
            *   45473    45572: gap of unknown length
```

```
            *    45573    47036: contig of 1464 bp in length
            *    47037    47136: gap of unknown length
            *    47137    48244: contig of 1108 bp in length
            *    48245    48344: gap of unknown length
            *    48345    50213: contig of 1869 bp in length
            *    50214    50313: gap of unknown length
            *    50314    51892: contig of 1579 bp in length
            *    51893    51992: gap of unknown length
            *    51993    53712: contig of 1720 bp in length
            *    53713    53812: gap of unknown length
            *    53813    55724: contig of 1912 bp in length
            *    55725    55824: gap of unknown length
            *    55825    58414: contig of 2590 bp in length
            *    58415    58514: gap of unknown length
            *    58515    60085: contig of 1571 bp in length
            *    60086    60185: gap of unknown length
            *    60186    62192: contig of 2007 bp in length
            *    62193    62292: gap of unknown length
            *    62293    64056: contig of 1764 bp in length
            *    64057    64156: gap of unknown length
            *    64157    66995: contig of 2839 bp in length
            *    66996    67095: gap of unknown length
            *    67096    69950: contig of 2855 bp in length
            *    69951    70050: gap of unknown length
            *    70051    71752: contig of 1702 bp in length
            *    71753    71852: gap of unknown length
            *    71853    73480: contig of 1628 bp in length
            *    73481    73580: gap of unknown length
            *    73581    75880: contig of 2300 bp in length
            *    75881    75980: gap of unknown length
            *    75981    77979: contig of 1999 bp in length
            *    77980    78079: gap of unknown length
            *    78080    79749: contig of 1670 bp in length
            *    79750    79849: gap of unknown length
            *    79850    82131: contig of 2282 bp in length
            *    82132    82231: gap of unknown length
            *    82232    84164: contig of 1933 bp in length
            *    84165    84264: gap of unknown length
            *    84265    86301: contig of 2037 bp in length
            *    86302    86401: gap of unknown length
            *    86402    89582: contig of 3181 bp in length

  Query Match              5.7%;  Score 23;  DB 2;  Length 173852;
  Best Local Similarity   100.0%;  Pred. No. 0.073;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    149 AGATGTCAAGGTCTTCACAGAGA 171
          |||||||||||||||||||||||
Db  91195 AGATGTCAAGGTCTTCACAGAGA 91217


RESULT 2
AC120750
LOCUS       AC120750              176340 bp    DNA     linear   HTG 24-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-54I13, *** SEQUENCING IN PROGRESS
            ***, 62 unordered pieces.
ACCESSION   AC120750
VERSION     AC120750.3  GI:21908122
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Rattus norvegicus.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 176340)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176340)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 176340)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 19, 2002 this sequence version replaced gi:20564372.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GXZE
                Center clone name: CH230-54I13
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 127817 bases at least Q40
                Consensus quality: 134179 bases at least Q30
                Consensus quality: 139091 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 62 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
```

```
*       1     1036: contig of 1036 bp in length
*    1037     1136: gap of unknown length
*    1137     2875: contig of 1739 bp in length
*    2876     2975: gap of unknown length
*    2976     4117: contig of 1142 bp in length
*    4118     4217: gap of unknown length
*    4218     5228: contig of 1011 bp in length
*    5229     5328: gap of unknown length
*    5329     6834: contig of 1506 bp in length
*    6835     6934: gap of unknown length
*    6935     8520: contig of 1586 bp in length
*    8521     8620: gap of unknown length
*    8621    10074: contig of 1454 bp in length
*   10075    10174: gap of unknown length
*   10175    11457: contig of 1283 bp in length
*   11458    11557: gap of unknown length
*   11558    12754: contig of 1197 bp in length
*   12755    12854: gap of unknown length
*   12855    14356: contig of 1502 bp in length
*   14357    14456: gap of unknown length
*   14457    15696: contig of 1240 bp in length
*   15697    15796: gap of unknown length
*   15797    17159: contig of 1363 bp in length
*   17160    17259: gap of unknown length
*   17260    18560: contig of 1301 bp in length
*   18561    18660: gap of unknown length
*   18661    19681: contig of 1021 bp in length
*   19682    19781: gap of unknown length
*   19782    22054: contig of 2273 bp in length
*   22055    22154: gap of unknown length
*   22155    23585: contig of 1431 bp in length
*   23586    23685: gap of unknown length
*   23686    25187: contig of 1502 bp in length
*   25188    25287: gap of unknown length
*   25288    26708: contig of 1421 bp in length
*   26709    26808: gap of unknown length
*   26809    27945: contig of 1137 bp in length
*   27946    28045: gap of unknown length
*   28046    29741: contig of 1696 bp in length
*   29742    29841: gap of unknown length
*   29842    31336: contig of 1495 bp in length
*   31337    31436: gap of unknown length
*   31437    33911: contig of 2475 bp in length
*   33912    34011: gap of unknown length
*   34012    35240: contig of 1229 bp in length
*   35241    35340: gap of unknown length
*   35341    36536: contig of 1196 bp in length
*   36537    36636: gap of unknown length
*   36637    38104: contig of 1468 bp in length
*   38105    38204: gap of unknown length
*   38205    40397: contig of 2193 bp in length
*   40398    40497: gap of unknown length
*   40498    42465: contig of 1968 bp in length
*   42466    42565: gap of unknown length
*   42566    43699: contig of 1134 bp in length
*   43700    43799: gap of unknown length
*   43800    45643: contig of 1844 bp in length
*   45644    45743: gap of unknown length
*   45744    47002: contig of 1259 bp in length
*   47003    47102: gap of unknown length
*   47103    49295: contig of 2193 bp in length
*   49296    49395: gap of unknown length
*   49396    51367: contig of 1972 bp in length
*   51368    51467: gap of unknown length
*   51468    53740: contig of 2273 bp in length
*   53741    53840: gap of unknown length
*   53841    56438: contig of 2598 bp in length
*   56439    56538: gap of unknown length
*   56539    59398: contig of 2860 bp in length
*   59399    59498: gap of unknown length
*   59499    62008: contig of 2510 bp in length
*   62009    62108: gap of unknown length
*   62109    63704: contig of 1596 bp in length
*   63705    63804: gap of unknown length
*   63805    65770: contig of 1966 bp in length
*   65771    65870: gap of unknown length
*   65871    69190: contig of 3320 bp in length
*   69191    69290: gap of unknown length
*   69291    71336: contig of 2046 bp in length
*   71337    71436: gap of unknown length
*   71437    75051: contig of 3615 bp in length
*   75052    75151: gap of unknown length
*   75152    77789: contig of 2638 bp in length
*   77790    77889: gap of unknown length
*   77890    79721: contig of 1832 bp in length
*   79722    79821: gap of unknown length
*   79822    83962: contig of 4141 bp in length
*   83963    84062: gap of unknown length
*   84063    87620: contig of 3558 bp in length
*   87621    87720: gap of unknown length
*   87721    90404: contig of 2684 bp in length
*   90405    90504: gap of unknown length
*   90505    94725: contig of 4221 bp in length
*   94726    94825: gap of unknown length
*   94826    98967: contig of 4142 bp in length
*   98968    99067: gap of unknown length
*   99068   102745: contig of 3678 bp in length
*  102746   102845: gap of unknown length
*  102846   106746: contig of 3901 bp in length
*  106747   106846: gap of unknown length
*  106847   110509: contig of 3663 bp in length
*  110510   110609: gap of unknown length
*  110610   114673: contig of 4064 bp in length
*  114674   114773: gap of unknown length
*  114774   119367: contig of 4594 bp in length
*  119368   119467: gap of unknown length
*  119468   123339: contig of 3872 bp in length

  Query Match              5.7%;  Score 23;  DB 2;  Length 176340;
  Best Local Similarity   100.0%;  Pred. No. 0.073;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     149 AGATGTCAAGGTCTTCACAGAGA 171
           |||||||||||||||||||||||
Db   47941 AGATGTCAAGGTCTTCACAGAGA 47963
```

RESULT 3
HUMCOL14/c

```
LOCUS       HUMCOL14                369 bp    DNA     linear   PRI 09-APR-1997
DEFINITION  Human 180 kDa bullous pemphigoid antigen 2/type XVII collagen
            (BPAG2/COL17A1) gene, exon 14.
ACCESSION   U76577
VERSION     U76577.1  GI:1825501
KEYWORDS    .
SEGMENT     14 of 41
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 369)
  AUTHORS   Gatalica,B., Pulkkinen,L., Li,K., Kuokkanen,K., Ryynanen,M.,
            McGrath,J.A. and Uitto,J.
  TITLE     Cloning of the human type XVII collagen gene (COL17A1), and
            detection of novel mutations in generalized atrophic benign
            epidermolysis bullosa
  JOURNAL   Am. J. Hum. Genet. 60 (2), 352-365 (1997)
  MEDLINE   97164601
   PUBMED   9012408
REFERENCE   2  (bases 1 to 369)
  AUTHORS   Gatalica,B., Pulkkinen,L., Li,K., Ryynanen,M., Kuokkanen,K.,
            McGrath,J.A. and Uitto,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-1996) Dermatology, Thomas Jefferson University,
            233 South 10th Street, Philadelphia, PA 19107, USA
FEATURES             Location/Qualifiers
```

```
     source          1. .369
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     intron          <1. .101
                     /gene="BPAG2/COL17A1"
                     /number=13
     exon            102. .263
                     /gene="BPAG2/COL17A1"
                     /number=14
     intron          264. .>369
                     /gene="BPAG2/COL17A1"
                     /number=14
BASE COUNT       82 a    110 c     76 g    101 t
ORIGIN

  Query Match             5.0%;  Score 20;  DB 9;  Length 369;
  Best Local Similarity   100.0%;  Pred. No. 5.1;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      40 GACATGTGCCAGGAGAAGGG 59
           ||||||||||||||||||||
Db      70 GACATGTGCCAGGAGAAGGG 51


RESULT 4
AF063186/c
LOCUS       AF063186               2964 bp    DNA     linear   BCT 02-JAN-2001
DEFINITION  Arthrobacter paraffineus DNA polymerase I gene, complete cds.
ACCESSION   AF063186
VERSION     AF063186.1  GI:12002009
KEYWORDS    .
SOURCE      Rhodococcus erythropolis.
  ORGANISM  Rhodococcus erythropolis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
REFERENCE   1  (bases 1 to 2964)
  AUTHORS   Huang,Y.P. and Ito,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-1998) Microbiology and Immunology, University of
            Arizona, 1501 N. Campbell Ave., Tucson, AZ 85724, USA
FEATURES             Location/Qualifiers
     source          1. .2964
                     /organism="Rhodococcus erythropolis"
                     /strain="ATCC19558"
                     /db_xref="ATCC:19558"
                     /db_xref="taxon:1833"
     CDS             112. .2874
                     /codon_start=1
                     /transl_table=11
                     /product="DNA polymerase I"
                     /protein_id="AAG43148.1"
                     /db_xref="GI:12002010"
                     /translation="MPPFVLVDGSYFLFRAFHALPPLTTSTGLHTNAIRGAISAIQKL
                     MRRTQPTHMAVIFDTPEPTFRHELSPIYKGDRPSMPTELSEQIPYLHALIRALGIPLH
                     TLPGAEADDIIGTLAKRAEAMGHQVLISTGDKDMAQLVTEKVTLEDSFKEKPLDVDGV
                     FEKFGVWPNQIIDYLTLMGDASDGIMGVPGVGAKTAAKLLTEYGSIGGILENVDKIKG
                     KVGQNIKEHVDGIALDHQLASIVIDLDLNFGYDDLKLGEPNIETLRHLYTELEFRNQL
                     QSLDHPNNPNNSNYQQAAKSINAKAVTAPEPLEDHASVTSSDDQLGQATYHTVLNQQD
                     WDTLFERLSTEKRFAFDTETTSLDYRIAQIVGFSVAFDAQDAYYVPLAHDYEGAPEQL
                     NREVILAQIKPILEDESVKKIGHHLKYDAHVLENHGIHLAGWYFDTMLASYVLYSVAT
                     RHGMDDVARLYLSHLTTTYEQVAGRGAKQKTFNQIPLETAAHYAAEDAHVTYRLYEVL
                     SSKLQAHPELVNILHNIEVPVARVLTQMEENGIELDLAFLDQLGGEFSNTMQNLENQI
                     MEIAGESFNVSSPKQVGEVLFEKLGLKGGKKTTTGQYSTSESVLEKIEHPIAQLILEY
                     RGLSKLKSTYTDGLCKQANPDTHRVHTSYHQALTATGRLSSTDPNLQNIPIRAEIGRQ
                     IRKAFVAPEGRVLLAADYSQIELRLMAHLSQDEALLDAFIHGQDVHRRTAAEVLGIPL
                     EEVTNDQRRQAKAVNFGLLYGMSEFGLIRQLGFTRQESQDYIKQYFHRYPGIYDYMQR
                     TRQVALEQGFVETLLGRRLYTPDIDARNMMVRKAAERAAINAPLQGSAADIIKMAMIE
                     VDKMLPKDQAKMLLQVHDELVFEVDEDIADELAPKLAEVMQSVLQISVPLVVEVGKGK
                     NWDEAH"
BASE COUNT      823 a    698 c    696 g    747 t
ORIGIN

  Query Match             5.0%;  Score 20;  DB 1;  Length 2964;
  Best Local Similarity   100.0%;  Pred. No. 4.9;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     315 CATAATGCCATCTGATGCAT 334
           ||||||||||||||||||||
Db     669 CATAATGCCATCTGATGCAT 650


RESULT 5
SPAC19G12
LOCUS       SPAC19G12              37000 bp    DNA     linear   PLN 12-OCT-1999
DEFINITION  S.pombe chromosome I cosmid c19G12.
ACCESSION   Z97209
VERSION     Z97209.1  GI:2879763
KEYWORDS    aldose reductase; carboxypeptidase Y; conserved hypothetical;
            cut20+; cyclosome/apc subunit cut20/apc4; dna mismatch repair
            protein; fatty acid hydroxylase; histone h2a-beta; hta2;
            Mitochondrial energy transfer protein; pcy1;
            phosphatidylinositol-4-phosphate; pms1; polysaccharide deacetylase;
            prenyl transferase; RNA-binding; RNP-1; splicing factor;
            trehalose-phosphatase; tricarboxylate transport; tRNA-Leu;
            ureidoglycolate hydrolase.
SOURCE      Schizosaccharomyces pombe.
  ORGANISM  Schizosaccharomyces pombe
            Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
            Schizosaccharomycetales; Schizosaccharomycetaceae;
            Schizosaccharomyces.
REFERENCE   1  (bases 1 to 37000)
  AUTHORS   Oliver,K., Harris,D., Wood,V., Barrell,B.G. and Rajandream,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUN-1997) Schizosaccharomyces pombe chromosome I
            sequencing project, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk
COMMENT     On Feb 13, 1998 this sequence version replaced gi:2239193.
            Notes:
            Details of yeast sequencing at the Sanger Centre are available on
            the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_pombe/)
            Protein coding regions (CDS) have been predicted with the help of
            computer analysis using the Genefinder program in PomBase (an ACEDB
            database) with additional predictions for the branch-acceptor sites
            supplied by the program Sp3splice. CAUTION: It is possible that for
            any individual CDS we may have underestimated or overestimated the
            number of introns/exons or we may not have chosen the correct
            splice donor/acceptor sites.
            CDS are numbered using the following system eg SPAC5H10.01c. SP (S.
            pombe), A (chromosome 1), c5H10 (cosmid name), .01 (first CDS), c
            (complementary strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous.
            The length in codons is given for each CDS.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone.  It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange for a
            small overlap between neighbouring submissions.
            Cosmid c19G12 is overlapped at the 3' end by cosmid c23A1.
FEATURES             Location/Qualifiers
     source          1. .37000
                     /organism="Schizosaccharomyces pombe"
                     /strain="972h-"
                     /db_xref="taxon:4896"
                     /chromosome="I"
                     /map="IR"
                     /clone="cosmid c19G12"
     gene            complement(join(1. .823,869. .926,1018. .1140,1183. .1402,
                     1451. .1688))
                     /gene="cut20+"
                     /note="SPAC19G12.01c"
     CDS             complement(join(1. .823,869. .926,1018. .1140,1183. .1402,
                     1451. .1688))
                     /partial
                     /gene="cut20+"
                     /note="SPAC19G12.01c, len:487"
```

```
                /codon_start=1
                /label=cut20+
                /product="cyclosome/apc subunit cut20/apc4"
                /protein_id="CAB53521.1"
                /db_xref="GI:5791550"
                /translation="MVSKSFKYPKKHKFEWINETNRVERVGSGLKRVILCPSMELIAI
                LTCSNHLICCRSNSQRIWDVDFHDLEATELCWNHDGNLIVVGFKNGELKIIDSSTGHL
                VEQRPASRDLAVLMITWAMQETIVNEKRNDFLFDATAYMPLLGTLPSSAKEERIFSSK
                AIAQFFEPRKREGENNKKVELLSILDERGIRYINMFSSYKIGESDSLKSALNLGVPIS
                HSITNDLAYHVLICKGGTNISLKTLYMPLLKNDLGSIVDIATMSTRMQHLVRYLEEVL
                NAMYEEFDNVFKSEASFIKTFDALVSKYSDTTFFSLQLELYQFIMNGIPSDLLKEWIN
                ERVGDRVLKNWERAMVNSYTSLIIFCQEFVIPACERLTVLLSSARGKSIWGHMKGNTL
                LDAKLVEDCLATLGYLQNNVFSFLNCLFEEKKYMKHFISWLNYAIVEFNTSEPSSIPP
                QEIIEHINETVIYIRHSLFRSKLTSYFMGTKPLQLRD"
gene            complement(join(2169. .2404,2446. .3702,3774. .4239,
                4342. .4767))
                /gene="pms1"
                /note="SPAC19G12.02c"
CDS             complement(join(2169. .2404,2446. .3702,3774. .4239,
                4342. .4767))
                /gene="pms1"
                /note="SPAC19G12.02c, len:794"
                /codon_start=1
                /product="DNA mismatch repair protein pms1"
                /protein_id="CAB10113.1"
                /db_xref="GI:2239195"
                /db_xref="SWISS-PROT:P54280"
                /translation="MSTVKPIDANTVHKICSGQVITDVASAVKELVENSLDSGATTIE
                IRFKNYGINSIEVVDNGSGIDAGDYESIGKKHFTSKITDFEDLEALQTFGFRGEALSS
                LCAVGQVIISTATQNEAPKGVQLNLDHEGSLKDKLTIPFQRGTSVMVNDLFCTLPVRR
                KLLEKNYKREFSKAISLLQAYATISTNKRFMVYHQTKNSGKLLQLSTNSNKDMKLNIM
                NVFGTKVSSSLIPWNDGIIEGYISRPHVGSTRASNERQMLFINRRLVNLPKIARVIQE
                VFKPYSMAQSPFFAINLRITNGTIDINVSPDKKSVFLSEEDSIIEFIKNSLQNLCESC
                GHAISCSRSQSIFSYSSQIPDSSGDSTDQELPQSIPATESETSDDSSFSYKRSPCKRK
                LVEATAQPAISTSVAEGASLAQVSKPLPERLQKDSMRRSSPLNEKVTASSERMKKKLA
                LFASSTDTSMQKTIDSSFPLKQPINKPSSNPNNLLLNDPSPASTPVAKTINLNEIESV
                HNAESVSTLSSIPRTEQTSVANRIPSKTAALQKLKFFQSRPLDGLNKFSKKINISLSG
                VQKDIVRSDALLKFSNKIGVVHDISDENQEDHLNLTVHKADFLRMRVVGQFNRGFIVV
                VHGNNLFIIDQHASDEKFNYEHLKSNLVINSQDLVLPKRLDLAATEETVLIDHIDLIR
                RKGFGVAIDLNQRVGNRCTLLSVPTSKNVIFDTSDLLEIISVLSEHPQIDPFSSRLER
                MLASKACRSSVMIGRALTISEMNTIVRHLAELSKPWNCPHGRPTMRHLLRLKDI"
misc_feature    complement(2405. .2417)
                /gene="pms1"
                /note="ttaacttctgtag, splice branch and acceptor"
misc_feature    complement(2440. .2445)
                /gene="pms1"
                /note="gtaagt, splice donor sequence"
misc_feature    complement(join(3623. .3702,3774. .4239,4342. .4758))
                /gene="pms1"
                /note="Match to PF01119 DNA_mis_repair, DNA mismatch
                repair protein Score 529.60"
misc_feature    complement(3703. .3719)
                /gene="pms1"
                /note="ctaaccaattattttag, splice branch and acceptor"
misc_feature    complement(3768. .3773)
                /gene="pms1"
                /note="gtatgt, splice donor sequence"
misc_feature    complement(4240. .4253)
                /gene="pms1"
                /note="ctaactgaatgtag, splice branch and acceptor"
misc_feature    complement(4336. .4341)
                /gene="pms1"
                /note="gtaagt, splice donor sequence"
misc_feature    complement(4468. .4488)
                /gene="pms1"
                /note="PS00058 DNA mismatch repair proteins mutL"
gene            5410. .6372
                /gene="SPAC19G12.03"
CDS             5410. .6372
                /gene="SPAC19G12.03"
                /note="SPAC19G12.03, len:320, LOW SIMILARITY:Yersinia
                pseudotuberculosis., CAB46604, hypothetical 33.7 kd
                protein., (293 aa), fasta scores: opt: 155, E():0.003,
                (27.5% identity in 265 aa)"
                /codon_start=1
                /label=SPAC19G12.03
                /product="possible polysaccharide deacetylase"
                /protein_id="CAB10114.1"
                /db_xref="GI:2239196"
                /db_xref="SPTREMBL:O13842"
                /translation="MYETRDLTGNAGKPVDTNPWPNNSKIAVSFVVNYEEGGERSLLY
                EDEGFETFLTEAGLMPFPNRPVRERSIESCFEYGSRCGFWRILNLFKKHKVPFTCWAI
                GQAVEKNPVVVGAMEEAGCEVGSHSHRWINYEGVPPETEYEHIKKSVQAIQKASPSNS
                APRSWYTGRASLNTRKLVCQVYKDLGLPQPFDSDEYNDDYPYWVADPLASKPGAEDDK
                GLLIVPYTLEVNDMKYAVAPGFCNSDDFYTYARDAFDVLYEEGLEGAPKMMTIGLHCR
                LTGRPGRFRGLQKLMEHITSKEGVWVATREQIAQAWSAKHPYKA"
misc_feature    5608. .6042
                /gene="SPAC19G12.03"
                /note="Pfam match to entry PF01522 Polysac_deacet"
gene            6888. .7463
                /gene="SPAC19G12.04"
CDS             6888. .7463
                /gene="SPAC19G12.04"
                /note="SPAC19G12.04, len:191, SIMILARITY:Saccharomyces
                cerevisiae, DAL3_YEAST, ureidoglycolate hydrolase, (195
                aa), fasta scores: opt: 360, E():5e-17, (36.5% identity in
                200 aa)"
                /codon_start=1
                /label=SPAC19G12.04
                /product="putative ureidoglycolate hydrolase"
                /protein_id="CAB10115.1"
                /db_xref="GI:2239197"
                /db_xref="SPTREMBL:O13843"
                /translation="MEAKKIYAQALTDEAFAPFGSVVQQKDDVKMVSANGGTAKKYLK
                VSESIQNYEKSSSASTRKGVWNFFSTHPSVHPANDEHAAFQISVLERHPFTTQTFIPM
                CRSSDEQAYLIAVAPNAPDGMPDWNQTQAFVAKGAQGVTYSAGVWHAPMVTIGKETML
                AAFNYENGVAEDDCQVQSTESPIEVFIKIST"
gene            8397. .9272
                /gene="SPAC19G12.05"
CDS             8397. .9272
                /gene="SPAC19G12.05"
                /note="SPAC19G12.05, len:291, SIMILARITY:Saccharomyces
                cerevisiae, TXTP_YEAST, tricarboxylate transport protein,
                (299 aa), fasta scores: opt: 738, E():0, (49.0% identity
                in 290 aa)"
                /codon_start=1
                /label=SPAC19G12.05
                /product="putative tricarboxylate transport protein"
                /protein_id="CAB10116.1"
                /db_xref="GI:2239198"
                /db_xref="SPTREMBL:O13844"
                /translation="MSTVAKTNPKSSNKPGPVKSIIAGGVAGAIEISITYPAEFAKTR
                LQLYRNVEGTKAKLPPFGLEWYRGCSTVIVGNSLKAAVRFFAFDSIKKSLSDEHGHLT
                GPRTVLAGLGAGVAESVLVLTPFESIKTAIIDDRKRPNPRLKGFLQASRIIVHENGIR
                GLYRGLAATVARQAANSGVRFTAYNSIKQSLQSRLPPDEKLSTVTTFLVGSVAGIITV
                YCTQPIDTVKSRMQSLSASKEYKNSIHCAYKILTQDGLLRFWSGATPRLARLILSGGI
                VFTVYEKVMEILKPF"
misc_feature    8397. .9020
                /gene="SPAC19G12.05"
                /note="Pfam match to entry mito_carr PF00153,
                Mitochondrial carrier proteins"
misc_feature    8505. .8531
                /gene="SPAC19G12.05"
                /note="PS00215 Mitochondrial energy transfer proteins
                signa ture"
gene            complement(10121. .10516)
                /gene="hta2"
```

```
  Query Match             5.0%;  Score 20;  DB 8;  Length 37000;
  Best Local Similarity   100.0%;  Pred. No. 4.6;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     339 TTTCTGGTTTCTTTTGAACA 358
           ||||||||||||||||||||
Db   15917 TTTCTGGTTTCTTTTGAACA 15936
```

```
RESULT 6
AC100210/c
LOCUS       AC100210               59237 bp    DNA     linear   HTG 22-NOV-2001
DEFINITION  Mus musculus clone RP23-59M9, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC100210
VERSION     AC100210.1  GI:17047576
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 59237)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-59M9
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 59237)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L14347
                Center clone name: 59_M_9
            --------------.
            * NOTE: This record contains 74 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      693: contig of 693 bp in length
            *      694 793: gap of      100 bp
            *      794     1492: contig of 699 bp in length
            *     1493 1592: gap of      100 bp
            *     1593     2294: contig of 702 bp in length
            *     2295 2394: gap of      100 bp
            *     2395     3040: contig of 646 bp in length
            *     3041 3140: gap of      100 bp
            *     3141     3850: contig of 710 bp in length
            *     3851 3950: gap of      100 bp
            *     3951     4664: contig of 714 bp in length
            *     4665 4764: gap of      100 bp
            *     4765     5477: contig of 713 bp in length
            *     5478 5577: gap of      100 bp
            *     5578     6291: contig of 714 bp in length
            *     6292 6391: gap of      100 bp
            *     6392     7107: contig of 716 bp in length
            *     7108 7207: gap of      100 bp
            *     7208     7906: contig of 699 bp in length
            *     7907 8006: gap of      100 bp
            *     8007     8704: contig of 698 bp in length
            *     8705 8804: gap of      100 bp
            *     8805     9477: contig of 673 bp in length
            *     9478 9577: gap of      100 bp
            *     9578    10241: contig of 664 bp in length
            *    10242 10341: gap of      100 bp
            *    10342    11047: contig of 706 bp in length
            *    11048 11147: gap of      100 bp
            *    11148    11839: contig of 692 bp in length
            *    11840 11939: gap of      100 bp
            *    11940    12638: contig of 699 bp in length
            *    12639 12738: gap of      100 bp
            *    12739    13446: contig of 708 bp in length
            *    13447 13546: gap of      100 bp
            *    13547    14237: contig of 691 bp in length
            *    14238 14337: gap of      100 bp
            *    14338    15022: contig of 685 bp in length
            *    15023 15122: gap of      100 bp
            *    15123    15813: contig of 691 bp in length
            *    15814 15913: gap of      100 bp
            *    15914    16617: contig of 704 bp in length
            *    16618 16717: gap of      100 bp
            *    16718    17394: contig of 677 bp in length
            *    17395 17494: gap of      100 bp
            *    17495    18188: contig of 694 bp in length
            *    18189 18288: gap of      100 bp
            *    18289    18966: contig of 678 bp in length
            *    18967 19066: gap of      100 bp
            *    19067    19749: contig of 683 bp in length
            *    19750 19849: gap of      100 bp
            *    19850    20566: contig of 717 bp in length
            *    20567 20666: gap of      100 bp
            *    20667    21375: contig of 709 bp in length
            *    21376 21475: gap of      100 bp
            *    21476    22181: contig of 706 bp in length
            *    22182 22281: gap of      100 bp
            *    22282    22984: contig of 703 bp in length
            *    22985 23084: gap of      100 bp
            *    23085    23755: contig of 671 bp in length
            *    23756 23855: gap of      100 bp
            *    23856    24562: contig of 707 bp in length
            *    24563 24662: gap of      100 bp
            *    24663    25385: contig of 723 bp in length
            *    25386 25485: gap of      100 bp
            *    25486    26192: contig of 707 bp in length
            *    26193 26292: gap of      100 bp
            *    26293    26995: contig of 703 bp in length
            *    26996 27095: gap of      100 bp
            *    27096    27799: contig of 704 bp in length
            *    27800 27899: gap of      100 bp
            *    27900    28616: contig of 717 bp in length
            *    28617 28716: gap of      100 bp
            *    28717    29419: contig of 703 bp in length
            *    29420 29519: gap of      100 bp
            *    29520    30212: contig of 693 bp in length
            *    30213 30312: gap of      100 bp
            *    30313    31000: contig of 688 bp in length
            *    31001 31100: gap of      100 bp
            *    31101    31835: contig of 735 bp in length
            *    31836 31935: gap of      100 bp
            *    31936    32635: contig of 700 bp in length
```

```
            *    32636 32735: gap of      100 bp
            *    32736    33445: contig of 710 bp in length
            *    33446 33545: gap of      100 bp
            *    33546    34256: contig of 711 bp in length
            *    34257 34356: gap of      100 bp
            *    34357    35057: contig of 701 bp in length
            *    35058 35157: gap of      100 bp
            *    35158    35865: contig of 708 bp in length
            *    35866 35965: gap of      100 bp
            *    35966    36670: contig of 705 bp in length
            *    36671 36770: gap of      100 bp
            *    36771    37576: contig of 806 bp in length
            *    37577 37676: gap of      100 bp
            *    37677    38391: contig of 715 bp in length
            *    38392 38491: gap of      100 bp
            *    38492    39208: contig of 717 bp in length
            *    39209 39308: gap of      100 bp
            *    39309    40017: contig of 709 bp in length
            *    40018 40117: gap of      100 bp
            *    40118    40827: contig of 710 bp in length
            *    40828 40927: gap of      100 bp
            *    40928    41612: contig of 685 bp in length
            *    41613 41712: gap of      100 bp
            *    41713    42406: contig of 694 bp in length
            *    42407 42506: gap of      100 bp
            *    42507    43203: contig of 697 bp in length
            *    43204 43303: gap of      100 bp
            *    43304    44000: contig of 697 bp in length
            *    44001 44100: gap of      100 bp
            *    44101    44807: contig of 707 bp in length
            *    44808 44907: gap of      100 bp
            *    44908    45615: contig of 708 bp in length
            *    45616 45715: gap of      100 bp
            *    45716    46420: contig of 705 bp in length
            *    46421 46520: gap of      100 bp
            *    46521    47227: contig of 707 bp in length
            *    47228 47327: gap of      100 bp
            *    47328    48044: contig of 717 bp in length
            *    48045 48144: gap of      100 bp
            *    48145    48846: contig of 702 bp in length
            *    48847 48946: gap of      100 bp
            *    48947    49652: contig of 706 bp in length
            *    49653 49752: gap of      100 bp
            *    49753    50447: contig of 695 bp in length
            *    50448 50547: gap of      100 bp
            *    50548    51225: contig of 678 bp in length
            *    51226 51325: gap of      100 bp
            *    51326    52026: contig of 701 bp in length
            *    52027 52126: gap of      100 bp
            *    52127    52844: contig of 718 bp in length
            *    52845 52944: gap of      100 bp
            *    52945    53649: contig of 705 bp in length
            *    53650 53749: gap of      100 bp
            *    53750    54463: contig of 714 bp in length
            *    54464 54563: gap of      100 bp
            *    54564    55259: contig of 696 bp in length
            *    55260 55359: gap of      100 bp

  Query Match              5.0%;  Score 20;  DB 2;  Length 59237;
  Best Local Similarity   100.0%;  Pred. No. 4.6;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      31 AGGCTTAAAGACATGTGCCA 50
           ||||||||||||||||||||
Db   23149 AGGCTTAAAGACATGTGCCA 23130


RESULT 7
AC119163/c
LOCUS       AC119163               60748 bp    DNA     linear   HTG 25-APR-2002
DEFINITION  Mus musculus clone RP24-527P22, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC119163
VERSION     AC119163.1  GI:20303744
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 60748)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-527P22
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 60748)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L19420
                Center clone name: 527_P_22
            --------------.
            * NOTE: This record contains 75 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      717: contig of 717 bp in length
            *      718 817: gap of      100 bp
            *      818     1532: contig of 715 bp in length
            *     1533 1632: gap of      100 bp
            *     1633     2353: contig of 721 bp in length
            *     2354 2453: gap of      100 bp
            *     2454     3155: contig of 702 bp in length
            *     3156 3255: gap of      100 bp
            *     3256     3961: contig of 706 bp in length
            *     3962 4061: gap of      100 bp
            *     4062     4782: contig of 721 bp in length
            *     4783 4882: gap of      100 bp
            *     4883     5589: contig of 707 bp in length
            *     5590 5689: gap of      100 bp
            *     5690     6395: contig of 706 bp in length
```

```
*     6396 6495: gap of      100 bp
*     6496     7213: contig of 718 bp in length
*     7214 7313: gap of      100 bp
*     7314     8030: contig of 717 bp in length
*     8031 8130: gap of      100 bp
*     8131     8860: contig of 730 bp in length
*     8861 8960: gap of      100 bp
*     8961     9679: contig of 719 bp in length
*     9680 9779: gap of      100 bp
*     9780    10515: contig of 736 bp in length
*    10516 10615: gap of      100 bp
*    10616    11349: contig of 734 bp in length
*    11350 11449: gap of      100 bp
*    11450    12165: contig of 716 bp in length
*    12166 12265: gap of      100 bp
*    12266    12978: contig of 713 bp in length
*    12979 13078: gap of      100 bp
*    13079    13772: contig of 694 bp in length
*    13773 13872: gap of      100 bp
*    13873    14587: contig of 715 bp in length
*    14588 14687: gap of      100 bp
*    14688    15433: contig of 746 bp in length
*    15434 15533: gap of      100 bp
*    15534    16238: contig of 705 bp in length
*    16239 16338: gap of      100 bp
*    16339    17065: contig of 727 bp in length
*    17066 17165: gap of      100 bp
*    17166    17871: contig of 706 bp in length
*    17872 17971: gap of      100 bp
*    17972    18655: contig of 684 bp in length
*    18656 18755: gap of      100 bp
*    18756    19450: contig of 695 bp in length
*    19451 19550: gap of      100 bp
*    19551    20277: contig of 727 bp in length
*    20278 20377: gap of      100 bp
*    20378    21094: contig of 717 bp in length
*    21095 21194: gap of      100 bp
*    21195    21738: contig of 544 bp in length
*    21739 21838: gap of      100 bp
*    21839    22533: contig of 695 bp in length
*    22534 22633: gap of      100 bp
*    22634    23350: contig of 717 bp in length
*    23351 23450: gap of      100 bp
*    23451    24172: contig of 722 bp in length
*    24173 24272: gap of      100 bp
*    24273    24969: contig of 697 bp in length
*    24970 25069: gap of      100 bp
*    25070    25785: contig of 716 bp in length
*    25786 25885: gap of      100 bp
*    25886    26596: contig of 711 bp in length
*    26597 26696: gap of      100 bp
*    26697    27419: contig of 723 bp in length
*    27420 27519: gap of      100 bp
*    27520    28247: contig of 728 bp in length
*    28248 28347: gap of      100 bp
*    28348    29064: contig of 717 bp in length
*    29065 29164: gap of      100 bp
*    29165    29875: contig of 711 bp in length
*    29876 29975: gap of      100 bp
*    29976    30686: contig of 711 bp in length
*    30687 30786: gap of      100 bp
*    30787    31500: contig of 714 bp in length
*    31501 31600: gap of      100 bp
*    31601    32302: contig of 702 bp in length
*    32303 32402: gap of      100 bp
*    32403    33107: contig of 705 bp in length
*    33108 33207: gap of      100 bp
*    33208    33919: contig of 712 bp in length
*    33920 34019: gap of      100 bp
*    34020    34732: contig of 713 bp in length
*    34733 34832: gap of      100 bp
*    34833    35558: contig of 726 bp in length
*    35559 35658: gap of      100 bp
*    35659    36382: contig of 724 bp in length
*    36383 36482: gap of      100 bp
*    36483    37195: contig of 713 bp in length
*    37196 37295: gap of      100 bp
*    37296    37992: contig of 697 bp in length
*    37993 38092: gap of      100 bp
*    38093    38796: contig of 704 bp in length
*    38797 38896: gap of      100 bp
*    38897    39620: contig of 724 bp in length
*    39621 39720: gap of      100 bp
*    39721    40428: contig of 708 bp in length
*    40429 40528: gap of      100 bp
*    40529    41251: contig of 723 bp in length
*    41252 41351: gap of      100 bp
*    41352    42071: contig of 720 bp in length
*    42072 42171: gap of      100 bp
*    42172    42908: contig of 737 bp in length
*    42909 43008: gap of      100 bp
*    43009    43740: contig of 732 bp in length
*    43741 43840: gap of      100 bp
*    43841    44553: contig of 713 bp in length
*    44554 44653: gap of      100 bp
*    44654    45363: contig of 710 bp in length
*    45364 45463: gap of      100 bp
*    45464    46172: contig of 709 bp in length
*    46173 46272: gap of      100 bp
*    46273    46996: contig of 724 bp in length
*    46997 47096: gap of      100 bp
*    47097    47794: contig of 698 bp in length
*    47795 47894: gap of      100 bp
*    47895    48591: contig of 697 bp in length
*    48592 48691: gap of      100 bp
*    48692    49393: contig of 702 bp in length
*    49394 49493: gap of      100 bp
*    49494    50209: contig of 716 bp in length
*    50210 50309: gap of      100 bp
*    50310    51007: contig of 698 bp in length
*    51008 51107: gap of      100 bp
*    51108    51792: contig of 685 bp in length
*    51793 51892: gap of      100 bp
*    51893    52609: contig of 717 bp in length
*    52610 52709: gap of      100 bp
*    52710    53434: contig of 725 bp in length
*    53435 53534: gap of      100 bp
*    53535    54240: contig of 706 bp in length
*    54241 54340: gap of      100 bp
*    54341    55065: contig of 725 bp in length
*    55066 55165: gap of      100 bp
*    55166    55877: contig of 712 bp in length
*    55878 55977: gap of      100 bp

  Query Match             5.0%;  Score 20;  DB 2;  Length 60748;
  Best Local Similarity   100.0%;  Pred. No. 4.6;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    225 TTGACTTCAGGTGGACAATT 244
          ||||||||||||||||||||
Db   4315 TTGACTTCAGGTGGACAATT 4296


RESULT 8
AC004149/c
LOCUS       AC004149               84767 bp    DNA     linear   PRI 06-JUN-2000
DEFINITION  Homo sapiens chromosome 17, clone hCIT.501_O_10, complete sequence.
ACCESSION   AC004149
VERSION     AC004149.1  GI:2981261
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 84767)
  AUTHORS   Birren,B., Fasman,K., McKernan,K., Nusbaum,C., Richardson,P. and
```

```
            Lander,E.
  TITLE     Homo sapiens chromosome 17, clone hCIT.501_O_10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 84767)
  AUTHORS   Birren,B., Fasman,K., McKernan,K., Nusbaum,C., Richardson,P.,
            Lander,E., Allen,N., Baker,J., Baldwin,J., Barna,N., Beckerly,R.,
            Boutwell,C., Byrne,S., Cantu,C., Castle,A., Cerny,J., Cooke,P.,
            Daly,M.J., Depayre,E., Devon,K., Dewar,K., Donelan,L., DuRette,B.,
            Etemadi,S., Ferreira,P., Forrest,C., Funke,R., Gage,D., Gardyna,S.,
            Gensheimer,S., Geraigery,K., Gilmartin,T., Gray,D., Hagos,B.,
            Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L., Linton,L.,
            MacKenzie,J., Marquis,N., McEwan,P., McGurk,A., Meldrim,J.,
            Molla,M., Morris,W., Morrow,J., Nachman,A., Naylor,J., O'Connor,T.,
            Pavlin,B., Peterson,K., Ranganath,S., Riley,R., Roberts,D.,
            Rollins,G., Rossello,R., Roy,A., Shyam,R., Soohoo,S.,
            Stange-Thomann,N., Stilwell,J., Stone,C., Strickland,C., Sydney,K.,
            Tang,L., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y.,
            Ye,W.J., Zemtseva,I., Zhao,J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 84767)
  AUTHORS   Birren,B., Fasman,K., McKernan,K., Nusbaum,C., Richardson,P.,
            Lander,E., Allen,N., Baker,J., Baldwin,J., Barna,N., Beckerly,R.,
            Benn,J., Boatin,C., Boutwell,C., Byrne,S., Cantu,C., Castle,A.,
            Cerny,J., Cooke,P., Daly,M.J., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., DuRette,B., Etemadi,S., Ferreira,P., FitzHugh,W.,
            Forrest,C., Funke,R., Gage,D., Gardyna,S., Gensheimer,S.,
            Geraigery,K., Gilmartin,T., Grant,G., Gray,D., Hagos,B., Harris,K.,
            Horton,L., Howland,J.C., Hui,L., Jacotot,L., Kann,L., Linton,L.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., Meldrim,J.,
            Molla,M., Morris,W., Morrow,J., Nachman,A., Nahf,R., Naylor,J.,
            O'Connor,T., Pavlin,B., Peterson,K., Riley,R., Roberts,D.,
            Rollins,G., Rossello,R., Roy,A., Shyam,R., Stange-Thomann,N.,
            Stilwell,J., Stone,C., Strickland,C., Subramanian,A., Sydney,K.,
            Tang,L., Vassiliev,H., Vo,A., Wagner,A., Wang,B., Wheeler,J.,
            Wu,Y., Ye,W.J., Zhao,J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAR-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 84767)
  AUTHORS   Birren,B., Fasman,K., McKernan,K., Nusbaum,C., Richardson,P.,
            Lander,E., Allen,N., Baker,J., Baldwin,J., Barna,N., Beckerly,R.,
            Benn,J., Boatin,C., Boutwell,C., Byrne,S., Cantu,C., Castle,A.,
            Cerny,J., Cooke,P., Daly,M.J., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., DuRette,B., Etemadi,S., Ferreira,P., FitzHugh,W.,
            Forrest,C., Funke,R., Gage,D., Gardyna,S., Gensheimer,S.,
            Geraigery,K., Gilmartin,T., Grant,G., Gray,D., Hagos,B., Harris,K.,
            Horton,L., Howland,J.C., Hui,L., Jacotot,L., Kann,L., Linton,L.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., Meldrim,J.,
            Molla,M., Morris,W., Morrow,J., Nachman,A., Nahf,R., Naylor,J.,
            O'Connor,T., Pavlin,B., Peterson,K., Riley,R., Roberts,D.,
            Rollins,G., Rossello,R., Roy,A., Shyam,R., Stange-Thomann,N.,
            Stilwell,J., Stone,C., Strickland,C., Subramanian,A., Sydney,K.,
            Tang,L., Vassiliev,H., Vo,A., Wagner,A., Wang,B., Wheeler,J.,
            Wu,Y., Ye,W.J., Zhao,J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 23, 1998 this sequence version replaced gi:2980974.
            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES             Location/Qualifiers
     source          1. .84767
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="hCIT.501_O_10"
                     /clone_lib="CITC Human BAC"
     repeat_region   150. .222
                     /rpt_family="LINE2"
     repeat_region   complement(289. .572)
                     /rpt_family="AluSx"
     repeat_region   complement(574. .707)
                     /rpt_family="AluSq/x"
     repeat_region   complement(2790. .2966)
                     /rpt_family="AluJb"
     repeat_region   complement(3027. .3322)
                     /rpt_family="AluY"
     repeat_region   complement(3400. .3509)
                     /rpt_family="MIR"
     repeat_region   3623. .3699
                     /rpt_family="(GGGA)n"
     repeat_region   4851. .4968
                     /rpt_family="MIR"
     repeat_region   5377. .5553
                     /rpt_family="MIR"
     repeat_region   complement(5725. .5910)
                     /rpt_family="MIR"
     repeat_region   complement(6877. .6906)
                     /rpt_family="(CA)n"
     repeat_region   complement(7333. .7379)
                     /rpt_family="MIR"
     repeat_region   7386. .7668
                     /rpt_family="AluY"
     repeat_region   complement(8737. .9039)
                     /rpt_family="AluJb"
     repeat_region   9141. .9206
                     /rpt_family="(CA)n"
     repeat_region   complement(9750. .10059)
                     /rpt_family="AluJb"
     repeat_region   10085. .10399
                     /rpt_family="MER58A"
     repeat_region   complement(10701. .10983)
                     /rpt_family="AluSp"
     repeat_region   complement(11001. .11172)
                     /rpt_family="AluSq"
     repeat_region   complement(11173. .11473)
                     /rpt_family="AluSg"
     repeat_region   complement(11476. .11553)
                     /rpt_family="AluSq"
     repeat_region   complement(11699. .11746)
                     /rpt_family="MER5B"
     repeat_region   complement(12281. .12397)
                     /rpt_family="L1MC4"
     repeat_region   12415. .12715
                     /rpt_family="AluSx"
     repeat_region   complement(12761. .12931)
                     /rpt_family="L1MD2"
     repeat_region   complement(12947. .13252)
                     /rpt_family="AluJb"
     repeat_region   complement(13253. .13622)
                     /rpt_family="L1ME3"
     repeat_region   13788. .14096
                     /rpt_family="AluSq"
     repeat_region   14109. .14353
                     /rpt_family="L1"
     repeat_region   14354. .14851
                     /rpt_family="MLT2CA"
     repeat_region   14852. .14990
                     /rpt_family="L1"
     repeat_region   14992. .15293
                     /rpt_family="AluSq"
     repeat_region   15307. .15899
                     /rpt_family="L1MB5"
     repeat_region   complement(15901. .16203)
                     /rpt_family="AluJo"
     repeat_region   complement(16409. .16499)
                     /rpt_family="L1MC3"
     repeat_region   complement(16503. .16802)
                     /rpt_family="AluSg"
     repeat_region   16959. .17266
                     /rpt_family="AluSx"
     repeat_region   18543. .18844
```

```
                     /rpt_family="AluSx"
     repeat_region   18845. .19131
                     /rpt_family="AluY"
     unsure          complement(19183. .19256)
                     /note="Single-stranded coverage here."
     repeat_region   complement(19245. .19359)
                     /rpt_family="AluY"
     repeat_region   complement(19509. .19679)
                     /rpt_family="MIR"
     repeat_region   20108. .20229
                     /rpt_family="MIR"
     repeat_region   20282. .20575
                     /rpt_family="AluJo"
     repeat_region   complement(20718. .20745)
                     /rpt_family="(GAA)n"
     repeat_region   complement(20746. .21048)
                     /rpt_family="AluSx"
     repeat_region   complement(21058. .21130)
                     /rpt_family="MIR"
     repeat_region   21640. .21749
                     /rpt_family="(CA)n"
     repeat_region   complement(22972. .23105)
                     /rpt_family="MIR"
     repeat_region   complement(23163. .23191)
                     /rpt_family="(CA)n"
     repeat_region   23191. .23247
                     /rpt_family="(GA)n"
     repeat_region   23359. .23411
                     /rpt_family="LINE2"
     repeat_region   complement(23516. .23605)
                     /rpt_family="MIR"
     repeat_region   complement(23623. .23732)
                     /rpt_family="(CA)n"
     repeat_region   complement(23800. .24071)
                     /rpt_family="AluSg"
     repeat_region   complement(25082. .25145)
                     /rpt_family="(TGGA)n"
     repeat_region   complement(25150. .25348)
                     /rpt_family="LINE2"
     repeat_region   complement(25396. .25685)
                     /rpt_family="AluSx"
     repeat_region   26709. .27007
                     /rpt_family="AluJo"

  Query Match             5.0%;  Score 20;  DB 9;  Length 84767;
  Best Local Similarity   100.0%;  Pred. No. 4.6;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    339 TTTCTGGTTTCTTTTGAACA 358
          ||||||||||||||||||||
Db  26525 TTTCTGGTTTCTTTTGAACA 26506


RESULT 9
AC091230/c
LOCUS       AC091230              128459 bp    DNA     linear   HTG 06-AUG-2002
DEFINITION  Homo sapiens chromosome 15 clone RP11-414J4 map 15, *** SEQUENCING
            IN PROGRESS ***, 4 unordered pieces.
ACCESSION   AC091230
VERSION     AC091230.9  GI:22123234
KEYWORDS    HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 128459)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-414J4
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 128459)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
            Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,
            Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
            Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 128459)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Aug 6, 2002 this sequence version replaced gi:20198593.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L11927
                Center clone name: 414_J_4
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    68585: contig of 68585 bp in length
            *    68586 68685: gap of      100 bp
            *    68686    80011: contig of 11326 bp in length
            *    80012 80111: gap of      100 bp
            *    80112   108353: contig of 28242 bp in length
            *   108354 108453: gap of      100 bp
            *   108454   128459: contig of 20006 bp in length.
FEATURES             Location/Qualifiers
     source          1. .128459
```

```
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-414J4"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    28709 a  33887 c  34863 g  30464 t    536 others
ORIGIN

  Query Match             5.0%;  Score 20;  DB 2;  Length 128459;
  Best Local Similarity   100.0%;  Pred. No. 4.5;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    131 GATGAGAAGAAGGAAAGGAG 150
          ||||||||||||||||||||
Db  50754 GATGAGAAGAAGGAAAGGAG 50735


RESULT 10
AL138761
LOCUS       AL138761              141970 bp    DNA     linear   PRI 28-JUL-2000
DEFINITION  Human DNA sequence from clone RP11-16H23 on chromosome 10. Contains
            the gene KIAA0204 (HSLK) for a protein kinase, the COL17A1 gene for
            collagen type XVII alpha 1 (BP180), ESTs and GSSs, complete
            sequence.
ACCESSION   AL138761
VERSION     AL138761.12  GI:8573811
KEYWORDS    HTG; BP180; COL17A1; collagen; HSLK; KIAA0204; protein kinase.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 141970)
  AUTHORS   Brown,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 20, 2000 this sequence version replaced gi:8452481.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            RP11-16H23 is from the library RPCI-11.1 constructed at the Roswell
            Park Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-16H23.
FEATURES             Location/Qualifiers
     source          1. .141970
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-16H23"
                     /clone_lib="RPCI-11.1"
     misc_feature    6. .478
                     /note="match: GSS: Em:B83625"
     repeat_region   41. .330
                     /note="L1ME1 repeat: matches 5870. .6161 of consensus"
     repeat_region   331. .506
                     /note="MER20 repeat: matches 9. .218 of consensus"
     repeat_region   768. .1052
                     /note="AluJo repeat: matches 1. .285 of consensus"
     repeat_region   2466. .2755
                     /note="AluSx repeat: matches 1. .289 of consensus"
     repeat_region   3051. .3567
                     /note="L1ME repeat: matches 4876. .5403 of consensus"
     repeat_region   3725. .5542
                     /note="L1MB3 repeat: matches 4371. .6182 of consensus"
     repeat_region   6046. .6250
                     /note="MER82 repeat: matches 446. .653 of consensus"
     repeat_region   6285. .6488
                     /note="MIR repeat: matches 2. .207 of consensus"
     repeat_region   6508. .6598
                     /note="MER82 repeat: matches 2. .92 of consensus"
     repeat_region   7560. .7611
                     /note="MIR repeat: matches 173. .224 of consensus"
     repeat_region   7764. .7857
                     /note="L2 repeat: matches 2456. .2548 of consensus"
     repeat_region   7858. .8152
                     /note="AluSx repeat: matches 1. .295 of consensus"
     repeat_region   8153. .8349
                     /note="L2 repeat: matches 2548. .2750 of consensus"
     repeat_region   8412. .8728
                     /note="MER44A repeat: matches 1. .333 of consensus"
     repeat_region   8988. .9065
                     /note="39 copies 2 mer tt 70% conserved"
     repeat_region   9528. .10072
                     /note="MER1A repeat: matches 1. .527 of consensus"
     repeat_region   10108. .10414
                     /note="AluSg repeat: matches 1. .308 of consensus"
     repeat_region   11314. .11808
                     /note="MER34 repeat: matches 8. .543 of consensus"
     repeat_region   11819. .11942
                     /note="MER85 repeat: matches 1. .135 of consensus"
     repeat_region   11956. .11985
                     /note="MER85 repeat: matches 1. .30 of consensus"
     repeat_region   12000. .12162
                     /note="MIR repeat: matches 23. .198 of consensus"
     repeat_region   12296. .12415
                     /note="MIR repeat: matches 18. .140 of consensus"
     repeat_region   12398. .12485
                     /note="L2 repeat: matches 2611. .2697 of consensus"
     repeat_region   12514. .12899
                     /note="L1MC1 repeat: matches 5186. .5561 of consensus"
     repeat_region   12900. .13206
                     /note="AluY repeat: matches 1. .307 of consensus"
     repeat_region   13207. .13912
                     /note="L1MC1 repeat: matches 5561. .6305 of consensus"
     repeat_region   13914. .13979
                     /note="L1M4 repeat: matches 5186. .5253 of consensus"
     gene            14114. .52672
                     /gene="bA16H23.1"
     mRNA            join(<14114. .14278,16037. .16085,16423. .16572,22338. .22410,
                     22558. .22752,23306. .23387,24883. .25011,25611. .26966,
                     29000. .29130,29351. .29474,31616. .31795,34255. .34347,
                     41599. .41728,42223. .42347,43173. .43361,43927. .44052,
                     45059. .45172,49013. .52672)
                     /gene="bA16H23.1"
                     /product="bA16H23.1.2 (protein kinase KIAA0204 (HSLK)
                     (isoform 2))"
                     /note="match: cDNAs: Em:D88425 Em:AB015718"
                     /evidence=not_experimental
     mRNA            join(<14114. .14278,16037. .16085,16423. .16572,22338. .22410,
                     22558. .22752,23306. .23387,24883. .25011,25611. .26966,
                     29000. .29130,29351. .29474,31616. .31795,41599. .41728,
                     42223. .42347,43173. .43361,43927. .44052,45059. .45172,
                     49013. .52672)
                     /gene="bA16H23.1"
```

```
                     /product="bA16H23.1.1 (protein kinase KIAA0204 (HSLK)
                     (isoform 1))"
                     /note="match: cDNAs: Em:AF112855 Em:AF039574 Em:D88425
                     Em:D86959 Em:AB003357 Em:AB002804
                     match: ESTs: Em:AA153777 Em:AA667275 Em:AI183019
                     Em:AW402772 Em:AA190321 Em:AA798865 Em:AL120836 Em:C18798
                     Em:AA880588 Em:AI048714 Em:AA717749 Em:AA268478 Em:H31623
                     Em:AA108424 Em:C05749 Em:AA491329 Em:D82765 Em:AA356218
                     Em:AW364095 Em:AW604227 Em:AW604224 Em:AW364079 Em:W17289
                     Em:AI424888 Em:AI559709 Em:D59390 Em:AA403026 Em:AA702324
                     Em:AW322697 Em:C15395 Em:AA362324 Em:AI074478 Em:AA128460
                     Em:AI123337 Em:Z45955 Em:AA147494 Em:AA398426 Em:AA772565
                     Em:AI338645 Em:AI378364 Em:AI142068 Em:R88093 Em:C14927
                     Em:AW024338 Em:AI350740 Em:AA621484 Em:D83865 Em:AW573553
                     Em:AA323083 Em:R17527 Em:T36207 Em:AI400071 Em:AA724517
                     Em:R78860 Em:AA493393 Em:AW300413 Em:AI039339 Em:AI758698
                     Em:F13095 Em:AI675770 Em:AA347410 Em:AW604225 Em:H44714
                     Em:AA307061 Em:H25828 Em:AW021959 Em:W24152 Em:H93694
                     Em:AA505137 Em:N54648 Em:N58456 Em:AI540244 Em:T75451
                     Em:AA454970 Em:AA479380 Em:AI206000 Em:AA644445
                     Em:AL134377 Em:C15275 Em:AI377832 Em:AI949218 Em:AI984184
                     Em:H09850 Em:AA922539 Em:AA367240 Em:AA292145 Em:AI367042
                     Em:N78544 Em:N91956 Em:N78186 Em:AI807630 Em:AA772218
                     Em:AA886401 Em:AA131987 Em:AI004269 Em:AA248981 Em:D60901
                     Em:AL046816 Em:H04093 Em:AI205999 Em:AI088423 Em:AA217520
                     Em:R12258 Em:R35086 Em:AA129911 Em:AI018333 Em:AA736544
                     Em:AA454971 Em:AI478214 Em:AI923383 Em:AI949538 Em:N21494
                     Em:T27333 Em:AA131682 Em:N78047 Em:AI421215 Em:T55950
                     Em:T39277 Em:N89756 Em:AL134564 Em:AI911945 Em:AI682543
                     Em:AI029638 Em:AA928756 Em:AI124895 Em:N52032 Em:AW325591
                     Em:AI662844 Em:AI051000 Em:AI613499"
                     /evidence=not_experimental
CDS                  join(<14114. .14278,16037. .16085,16423. .16572,22338. .22410,
                     22558. .22752,23306. .23387,24883. .25011,25611. .26966,
                     29000. .29130,29351. .29474,31616. .31795,34255. .34347,
                     41599. .41728,42223. .42347,43173. .43361,43927. .44052,
                     45059. .45172,49013. .49159)
                     /gene="bA16H23.1"
                     /note="isoform based on homology to Guinea pig
                     serine/threonine kinase (Tr:O55092).
                     match: proteins: Tr:O55092"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="bA16H23.1.2 (protein kinase KIAA0204 (HSLK)
                     (isoform 2))"
                     /protein_id="CAC00587.1"
                     /db_xref="GI:9588136"
                     /translation="AQNKETSVLAAAKVIDTKSEEELEDYMVEIDILASCDHPNIVKL
                     LDAFYYENNLWILIEFCAGGAVDAVMLELERPLTESQIQVVCKQTLDALNYLHDNKII
                     HRDLKAGNILFTLDGDIKLADFGVSAKNTRTIQRRDSFIGTPYWMAPEVVMCETSKDR
                     PYDYKADVWSLGITLIEMAEIEPPHHELNPMRVLLKIAKSEPPTLAQPSRWSSNFKDF
                     LKKCLEKNVDARWTTSQLLQHPFVTVDSNKPIRELIAEAKAEVTEEVEDGKEEDEEEE
                     TENSLPIPASKRASSDLSIASSEEDKLSQNACILESVSEKTERSNSEDKLNSKILNEK
                     PTTDEPEKAVEDINEHITDAQLEAMTELHDRTAVIKENEREKRPKLENLPDTEDQETV
                     DINSVSEGKENNIMITLETNIEHNLKSEEEKDQEKQQMFENKLIKSEEIKDTILQTVD
                     LVSQETGEKEANIQAVDSEVGLTKEDTQEKLGEDDKTQKDVISNTSDVIGTCEAADVA
                     QKVDEDSAEDTQSNDGKEVVEVGQKLINKPMVGPEAGGTKEVPIKEIVEMNEIEEGKN
                     KEQAINSSENIMDINEEPGTTEGEEITESSSTEEMEVRSVVADTDQKALGSEVQDASK
                     VTTQIDKEKKEIPVSIKKEPEVTVVSQPTEPQPVLIPSININSDSGENKEEIGSLSKT
                     ETILPPESENPKENDNDSGTGSTADTSSIDLNLSISSFLSKTKDSGSISLQETRRQKK
                     TLKKTRKFIVDGVEVSVTTSKIVTDSDSKTEELRFLRRQELRELRFLQKEEQRAQQQL
                     NSKLQQQREQIFRRFEQEMMSKKRQYDQEIENLEKQQKQTIERLEQEHTNRLRDEAKR
                     IKGEQEKELSKFQNMLKNRKKEVINEVEKAPKELRKELMKRRKEELAQSQHAQEQEFV
                     QKQQQELDGSLKKIIQQQKAELANIERECLNNKQQLMRAREAAIWELEERHLQEKHQL
                     LKQQLKDQYFMQRHQLLKRHEKETEQMQRYNQRLIEELKNRQTQERARLPKIQRSEAK
                     TRMAMFKKSLRINSTATPDQDRDKIKQFAAQEEKRQKNERMAQHQKHENQMRDLQLQC
                     EANVRELHQLQNEKCHLLVEHETQKLKELDEEHSQELKEWREKLRPRKKTLEEEFARK
                     LQEQEVFFKMTGESECLNPSTQSRISKFYPIPSLHSTGS"
CDS                  join(<14114. .14278,16037. .16085,16423. .16572,22338. .22410,
                     22558. .22752,23306. .23387,24883. .25011,25611. .26966,
                     29000. .29130,29351. .29474,31616. .31795,41599. .41728,
                     42223. .42347,43173. .43361,43927. .44052,45059. .45172,
                     49013. .49159)
                     /gene="bA16H23.1"

  Query Match          5.0%;  Score 20;  DB 9;  Length 141970;
  Best Local Similarity  100.0%;  Pred. No. 4.5;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      40 GACATGTGCCAGGAGAAGGG 59
           ||||||||||||||||||||
Db   83751 GACATGTGCCAGGAGAAGGG 83770


RESULT 11
AP001883/c
LOCUS       AP001883              143620 bp    DNA     linear   PRI 06-MAR-2001
DEFINITION  Homo sapiens genomic DNA, chromosome 11q, clone:RP11-140O23,
            complete sequence.
ACCESSION   AP001883
VERSION     AP001883.5  GI:13359355
KEYWORDS    HTG.
SOURCE      Homo sapiens DNA, clone:RP11-140O23.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in DataBase (2000)
REFERENCE   2  (bases 1 to 143620)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-APR-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On Mar 16, 2001 this sequence version replaced gi:12381916.
FEATURES             Location/Qualifiers
     source          1. .143620
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-140O23"
BASE COUNT    41715 a  28357 c  29648 g  43900 t
ORIGIN

  Query Match          5.0%;  Score 20;  DB 9;  Length 143620;
  Best Local Similarity  100.0%;  Pred. No. 4.5;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     292 GTGTTAATTGTGGAGGTTCA 311
           ||||||||||||||||||||
Db 122317 GTGTTAATTGTGGAGGTTCA 122298


RESULT 12
AC015690/c
LOCUS       AC015690              145695 bp    DNA     linear   HTG 13-JUL-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-34A20 map 11, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC015690
VERSION     AC015690.3  GI:9107982
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 145695)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-34A20
  JOURNAL   Unpublished
```

```
REFERENCE   2  (bases 1 to 145695)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 13, 2000 this sequence version replaced gi:6910854.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L1347
                Center clone name: 34_A_20
            --------------.
            * NOTE: This record contains 163 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      760: contig of 760 bp in length
            *      761 860: gap of      100 bp
            *      861     1636: contig of 776 bp in length
            *     1637 1736: gap of      100 bp
            *     1737     2494: contig of 758 bp in length
            *     2495 2594: gap of      100 bp
            *     2595     3361: contig of 767 bp in length
            *     3362 3461: gap of      100 bp
            *     3462     4246: contig of 785 bp in length
            *     4247 4346: gap of      100 bp
            *     4347     5140: contig of 794 bp in length
            *     5141 5240: gap of      100 bp
            *     5241     6025: contig of 785 bp in length
            *     6026 6125: gap of      100 bp
            *     6126     6928: contig of 803 bp in length
            *     6929 7028: gap of      100 bp
            *     7029     7788: contig of 760 bp in length
            *     7789 7888: gap of      100 bp
            *     7889     8671: contig of 783 bp in length
            *     8672 8771: gap of      100 bp
            *     8772     9531: contig of 760 bp in length
            *     9532 9631: gap of      100 bp
            *     9632    10429: contig of 798 bp in length
            *    10430 10529: gap of      100 bp
            *    10530    11304: contig of 775 bp in length
            *    11305 11404: gap of      100 bp
            *    11405    12183: contig of 779 bp in length
            *    12184 12283: gap of      100 bp
            *    12284    13075: contig of 792 bp in length
            *    13076 13175: gap of      100 bp
            *    13176    13973: contig of 798 bp in length
            *    13974 14073: gap of      100 bp
            *    14074    14866: contig of 793 bp in length
            *    14867 14966: gap of      100 bp
            *    14967    15748: contig of 782 bp in length
            *    15749 15848: gap of      100 bp
            *    15849    16628: contig of 780 bp in length
            *    16629 16728: gap of      100 bp
            *    16729    17512: contig of 784 bp in length
            *    17513 17612: gap of      100 bp
            *    17613    18443: contig of 831 bp in length
            *    18444 18543: gap of      100 bp
            *    18544    19306: contig of 763 bp in length
            *    19307 19406: gap of      100 bp
            *    19407    20176: contig of 770 bp in length
            *    20177 20276: gap of      100 bp
            *    20277    21055: contig of 779 bp in length
            *    21056 21155: gap of      100 bp
            *    21156    21932: contig of 777 bp in length
            *    21933 22032: gap of      100 bp
            *    22033    22812: contig of 780 bp in length
            *    22813 22912: gap of      100 bp
            *    22913    23702: contig of 790 bp in length
            *    23703 23802: gap of      100 bp
            *    23803    24592: contig of 790 bp in length
            *    24593 24692: gap of      100 bp
            *    24693    25487: contig of 795 bp in length
            *    25488 25587: gap of      100 bp
            *    25588    26377: contig of 790 bp in length
            *    26378 26477: gap of      100 bp
            *    26478    27289: contig of 812 bp in length
            *    27290 27389: gap of      100 bp
            *    27390    28171: contig of 782 bp in length
            *    28172 28271: gap of      100 bp
            *    28272    29053: contig of 782 bp in length
            *    29054 29153: gap of      100 bp
            *    29154    29937: contig of 784 bp in length
            *    29938 30037: gap of      100 bp
            *    30038    30815: contig of 778 bp in length
            *    30816 30915: gap of      100 bp
            *    30916    31688: contig of 773 bp in length
            *    31689 31788: gap of      100 bp
            *    31789    32555: contig of 767 bp in length
            *    32556 32655: gap of      100 bp
            *    32656    33437: contig of 782 bp in length
            *    33438 33537: gap of      100 bp
            *    33538    34337: contig of 800 bp in length
            *    34338 34437: gap of      100 bp
            *    34438    35186: contig of 749 bp in length
            *    35187 35286: gap of      100 bp
            *    35287    36083: contig of 797 bp in length
            *    36084 36183: gap of      100 bp
            *    36184    36967: contig of 784 bp in length
            *    36968 37067: gap of      100 bp
            *    37068    37852: contig of 785 bp in length
            *    37853 37952: gap of      100 bp
            *    37953    38723: contig of 771 bp in length
            *    38724 38823: gap of      100 bp
            *    38824    39574: contig of 751 bp in length
            *    39575 39674: gap of      100 bp
            *    39675    40456: contig of 782 bp in length
            *    40457 40556: gap of      100 bp
            *    40557    41319: contig of 763 bp in length
            *    41320 41419: gap of      100 bp
            *    41420    42175: contig of 756 bp in length
            *    42176 42275: gap of      100 bp
            *    42276    43048: contig of 773 bp in length
            *    43049 43148: gap of      100 bp
            *    43149    43907: contig of 759 bp in length
            *    43908 44007: gap of      100 bp
            *    44008    44805: contig of 798 bp in length
            *    44806 44905: gap of      100 bp
            *    44906    45698: contig of 793 bp in length
            *    45699 45798: gap of      100 bp
```

```
            *    45799    46576: contig of 778 bp in length
            *    46577 46676: gap of      100 bp
            *    46677    47448: contig of 772 bp in length
            *    47449 47548: gap of      100 bp
            *    47549    48346: contig of 798 bp in length
            *    48347 48446: gap of      100 bp
            *    48447    49229: contig of 783 bp in length
            *    49230 49329: gap of      100 bp
            *    49330    50099: contig of 770 bp in length
            *    50100 50199: gap of      100 bp
            *    50200    50986: contig of 787 bp in length
            *    50987 51086: gap of      100 bp
            *    51087    51879: contig of 793 bp in length
            *    51880 51979: gap of      100 bp
            *    51980    52758: contig of 779 bp in length
            *    52759 52858: gap of      100 bp
            *    52859    53636: contig of 778 bp in length
            *    53637 53736: gap of      100 bp
            *    53737    54600: contig of 864 bp in length
            *    54601 54700: gap of      100 bp
            *    54701    55483: contig of 783 bp in length
            *    55484 55583: gap of      100 bp
            *    55584    56347: contig of 764 bp in length
            *    56348 56447: gap of      100 bp
            *    56448    57239: contig of 792 bp in length
            *    57240 57339: gap of      100 bp
            *    57340    58123: contig of 784 bp in length
            *    58124 58223: gap of      100 bp
            *    58224    59003: contig of 780 bp in length
            *    59004 59103: gap of      100 bp
            *    59104    59868: contig of 765 bp in length
            *    59869 59968: gap of      100 bp
            *    59969    60761: contig of 793 bp in length
            *    60762 60861: gap of      100 bp
            *    60862    61658: contig of 797 bp in length
            *    61659 61758: gap of      100 bp
            *    61759    62531: contig of 773 bp in length
            *    62532 62631: gap of      100 bp
            *    62632    63427: contig of 796 bp in length
            *    63428 63527: gap of      100 bp

  Query Match              5.0%;  Score 20;  DB 2;  Length 145695;
  Best Local Similarity   100.0%;  Pred. No. 4.5;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     292 GTGTTAATTGTGGAGGTTCA 311
           ||||||||||||||||||||
Db  107053 GTGTTAATTGTGGAGGTTCA 107034


RESULT 13
AC068675
LOCUS       AC068675              145870 bp    DNA     linear   HTG 13-AUG-2002
DEFINITION  Homo sapiens chromosome 17 clone RP11-392O1 map 17, *** SEQUENCING
            IN PROGRESS ***, 4 unordered pieces.
ACCESSION   AC068675
VERSION     AC068675.10  GI:22123057
KEYWORDS    HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 145870)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-392O1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 145870)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 145870)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Aug 6, 2002 this sequence version replaced gi:18543015.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L10324
                Center clone name: 392_O_1
            ---------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    51255: contig of 51255 bp in length
            *    51256 51355: gap of      100 bp
            *    51356    55066: contig of 3711 bp in length
            *    55067 55166: gap of      100 bp
            *    55167    70769: contig of 15603 bp in length
            *    70770 70869: gap of      100 bp
            *    70870   145870: contig of 75001 bp in length.
FEATURES             Location/Qualifiers
     source          1. .145870
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
```

```
                     /map="17"
                     /clone="RP11-392O1"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    37936 a  35909 c  35452 g  36198 t    375 others
ORIGIN

  Query Match             5.0%;  Score 20;  DB 2;  Length 145870;
  Best Local Similarity   100.0%;  Pred. No. 4.5;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     339 TTTCTGGTTTCTTTTGAACA 358
           ||||||||||||||||||||
Db   21888 TTTCTGGTTTCTTTTGAACA 21907


RESULT 14
AP001980/c
LOCUS       AP001980               151937 bp    DNA     linear   PRI 06-MAR-2001
DEFINITION  Homo sapiens genomic DNA, chromosome 11q, clone:RP11-264L21,
            complete sequence.
ACCESSION   AP001980
VERSION     AP001980.5  GI:13359356
KEYWORDS    HTG.
SOURCE      Homo sapiens DNA, clone:RP11-264L21.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in DataBase (2000)
REFERENCE   2  (bases 1 to 151937)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAY-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On Mar 16, 2001 this sequence version replaced gi:12381923.
FEATURES             Location/Qualifiers
     source          1. .151937
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-264L21"
BASE COUNT    49369 a  28003 c  27962 g  46603 t
ORIGIN

  Query Match             5.0%;  Score 20;  DB 9;  Length 151937;
  Best Local Similarity   100.0%;  Pred. No. 4.5;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     292 GTGTTAATTGTGGAGGTTCA 311
           ||||||||||||||||||||
Db   10098 GTGTTAATTGTGGAGGTTCA 10079


RESULT 15
AC036150
LOCUS       AC036150               156186 bp    DNA     linear   HTG 24-AUG-2002
DEFINITION  Homo sapiens chromosome 11 clone RP11-264F13 map 11, WORKING DRAFT
            SEQUENCE, 19 unordered pieces.
ACCESSION   AC036150
VERSION     AC036150.2  GI:7705190
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 156186)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-264F13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 156186)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 156186)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 4, 2000 this sequence version replaced gi:7523818.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L9050
                Center clone name: 264_F_13
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 147950 bases at least Q40
                Consensus quality: 151905 bases at least Q30
                Consensus quality: 153513 bases at least Q20
```

```
            Insert size: 163000; agarose-fp
            Insert size: 154386; sum-of-contigs
            Quality coverage: 4.4 in Q20 bases; agarose-fp
            Quality coverage: 4.7 in Q20 bases; sum-of-contigs
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 19 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1      547: contig of 547 bp in length
          *      548 647: gap of      100 bp
          *      648     3155: contig of 2508 bp in length
          *     3156 3255: gap of      100 bp
          *     3256     6384: contig of 3129 bp in length
          *     6385 6484: gap of      100 bp
          *     6485     9217: contig of 2733 bp in length
          *     9218 9317: gap of      100 bp
          *     9318    13050: contig of 3733 bp in length
          *    13051 13150: gap of      100 bp
          *    13151    16192: contig of 3042 bp in length
          *    16193 16292: gap of      100 bp
          *    16293    20043: contig of 3751 bp in length
          *    20044 20143: gap of      100 bp
          *    20144    24399: contig of 4256 bp in length
          *    24400 24499: gap of      100 bp
          *    24500    30398: contig of 5899 bp in length
          *    30399 30498: gap of      100 bp
          *    30499    37678: contig of 7180 bp in length
          *    37679 37778: gap of      100 bp
          *    37779    45560: contig of 7782 bp in length
          *    45561 45660: gap of      100 bp
          *    45661    53184: contig of 7524 bp in length
          *    53185 53284: gap of      100 bp
          *    53285    63459: contig of 10175 bp in length
          *    63460 63559: gap of      100 bp
          *    63560    72171: contig of 8612 bp in length
          *    72172 72271: gap of      100 bp
          *    72272    83976: contig of 11705 bp in length
          *    83977 84076: gap of      100 bp
          *    84077    98526: contig of 14450 bp in length
          *    98527 98626: gap of      100 bp
          *    98627   115880: contig of 17254 bp in length
          *   115881 115980: gap of      100 bp
          *   115981   134149: contig of 18169 bp in length
          *   134150 134249: gap of      100 bp
          *   134250   156186: contig of 21937 bp in length.
FEATURES             Location/Qualifiers
     source          1. .156186
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-264F13"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .547
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    648. .3155
                     /note="assembly_fragment"
     misc_feature    3256. .6384
                     /note="assembly_fragment"
     misc_feature    6485. .9217
                     /note="assembly_fragment"
     misc_feature    9318. .13050
                     /note="assembly_fragment"
     misc_feature    13151. .16192
                     /note="assembly_fragment"
     misc_feature    16293. .20043
                     /note="assembly_fragment"
     misc_feature    20144. .24399
                     /note="assembly_fragment"
     misc_feature    24500. .30398
                     /note="assembly_fragment"
     misc_feature    30499. .37678
                     /note="assembly_fragment"
     misc_feature    37779. .45560
                     /note="assembly_fragment"
     misc_feature    45661. .53184
                     /note="assembly_fragment"
     misc_feature    53285. .63459
                     /note="assembly_fragment"
     misc_feature    63560. .72171
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    72272. .83976
                     /note="assembly_fragment"
     misc_feature    84077. .98526
                     /note="assembly_fragment"
     misc_feature    98627. .115880
                     /note="assembly_fragment"
     misc_feature    115981. .134149
                     /note="assembly_fragment"
     misc_feature    134250. .156186
                     /note="assembly_fragment"
BASE COUNT    46555 a  30845 c  30411 g  46571 t   1804 others
ORIGIN

  Query Match             5.0%;  Score 20;  DB 2;  Length 156186;
  Best Local Similarity   100.0%;  Pred. No. 4.5;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     292 GTGTTAATTGTGGAGGTTCA 311
           ||||||||||||||||||||
Db 105256 GTGTTAATTGTGGAGGTTCA 105275


Search completed: January  2, 2003, 16:08:34
Job time : 3743 secs
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: January 2, 2003, 15:01:56 ; Search time 357 Seconds
(without alignments)
2535.864 Million cell updates/sec

Title: US-09-850-147-1
Perfect score: 402
Sequence: 1 ccgggccgggctcaccagct.....ggcataacttcctgtgttcc 402

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	19	4.7	324	22	AAS38626	Novel human diagno
	2	19	4.7	615	24	ABN64583	Human cancer relat
	3	19	4.7	1444	21	AAC46559	Zea mays DNA fragm
c	4	19	4.7	2016	23	AAS69866	DNA encoding novel
	5	19	4.7	2020	23	AAS82153	DNA encoding novel
	6	18	4.5	360	20	AAV89959	EST clone CS765.
	7	18	4.5	872	21	AAC99130	Human pancreatic c
	8	18	4.5	899	22	AAH06889	Human cDNA clone (
	9	18	4.5	1149	21	AAA51824	Human melanocortin
c	10	18	4.5	1205	22	AAH14264	Human cDNA sequenc
c	11	18	4.5	1984	21	AAC69571	Human secreted pro
	12	18	4.5	3099	23	AAS73815	DNA encoding novel
	13	18	4.5	3099	23	AAS84712	DNA encoding novel
	14	18	4.5	3137	23	AAS72090	DNA encoding novel
	15	18	4.5	3288	22	AAH14592	Human cDNA sequenc
	16	18	4.5	4623	24	ABK09751	Human ovarian tumo
	17	18	4.5	4633	23	AAS69913	DNA encoding novel
	18	18	4.5	4689	23	AAS73822	DNA encoding novel
	19	18	4.5	5136	23	AAS84715	DNA encoding novel
c	20	18	4.5	25277	22	AAS30021	Human lung antigen
c	21	18	4.5	25277	22	AAS28714	Genomic sequence #
c	22	18	4.5	25279	22	AAS30020	Human lung antigen
c	23	18	4.5	25279	22	AAS28713	Genomic sequence #
c	24	18	4.5	30610	22	ABA15643	Human nervous syst
c	25	18	4.5	319608	21	AAH51601	Human chromosome 1
c	26	18	4.5	319608	22	AAS09301	Human schizophreni
	27	17	4.2	179	24	ABN96499	Gene #2997 used to
	28	17	4.2	179	24	ABL66186	Lung cancer relate
	29	17	4.2	203	14	AAQ60752	Human brain Expres
c	30	17	4.2	254	21	AAC16115	Human secreted pro
c	31	17	4.2	261	21	AAC10644	Human secreted pro
	32	17	4.2	320	14	AAQ60849	Human brain Expres
	33	17	4.2	324	20	AAX40512	Human secreted pro
	34	17	4.2	324	21	AAC29123	Human secreted pro
	35	17	4.2	327	20	AAH86650	Human single nucle
	36	17	4.2	327	20	AAH87393	Human single nucle
	37	17	4.2	370	24	ABN20928	Human ORFX polynuc
	38	17	4.2	378	20	AAV87181	EST clone BD489.
	39	17	4.2	462	22	AAH08928	Human cDNA clone (
	40	17	4.2	471	22	AAH52447	S. epidermidis ope
	41	17	4.2	471	22	AAH53375	S. epidermidis ope
c	42	17	4.2	488	22	ABA13256	Human nervous syst
	43	17	4.2	489	24	ABN90603	Staphylococcus epi
c	44	17	4.2	498	21	AAC99118	Human pancreatic c
	45	17	4.2	512	23	ABV50321	Human prostate exp

ALIGNMENTS

RESULT 1
AAS38626
ID AAS38626 standard; cDNA; 324 BP.
XX
AC AAS38626;
XX
DT 17-DEC-2001 (first entry)
XX
DE Novel human diagnostic and therapeutic gene #1684.
XX
KW Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.
XX
OS Homo sapiens.
XX
PN WO200166753-A2.
XX
PD 13-SEP-2001.
XX
PF 09-MAR-2001; 2001WO-US07787.
XX
PR 09-MAR-2000; 2000US-0188609.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Kassam A, Lamson G;
PI Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;
XX
DR WPI; 2001-530177/58.
XX

PT New polynucleotides and polypeptides, useful for diagnosis and
PT treatment of breast, lung and colon cancer -
XX
PS Claim 1; Page 1018; 1193pp; English.
XX
CC The invention relates to new polynucleotides and polypeptides, useful for
CC diagnosis and treatment of breast, lung and colon cancer. The sequences
CC can be used in detecting differentially expressed genes correlated with a
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample derived from a
CC cell suspected of being cancerous. They can also be used to inhibit
CC tumour growth by modulating expression of a gene product. AAS36943-
CC AAS39338 represent novel human diagnostic and therapeutic coding
CC sequences of the invention.
XX
SQ Sequence 324 BP; 86 A; 59 C; 74 G; 105 T; 0 other;

Query Match 4.7%; Score 19; DB 22; Length 324;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 346 TTTCTTTTGAACACAAGTA 364
||||||||||||||||||
Db 205 TTTCTTTTGAACACAAGTA 223

RESULT 2
ABN64583
ID ABN64583 standard; cDNA; 615 BP.
XX
AC ABN64583;
XX
DT 28-JUN-2002 (first entry)
XX
DE Human cancer related polynucleotide SEQ ID NO 4550.
XX
KW Human; cytostatic; gene expression; gene mapping; tissue profiling;
KW gene therapy; cancer; tumour; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200214500-A2.
XX
PD 21-FEB-2002.
XX
PF 16-AUG-2001; 2001WO-US25840.
XX
PR 16-AUG-2000; 2000US-226326P.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Escobedo J, Garcia PD, Sudduth-Klinger J, Reinhard C, Randazzo F;
PI Lamson G, Scott EM, Zhang G, Kassam A, Pot D, Labat I;
XX
DR WPI; 2002-241905/29.
XX
PT New nucleic acid for producing a polypeptide, detecting differentially
PT expressed genes correlated with a cancerous state of a mammalian cell,
PT and inhibiting tumor growth -
XX
PS Claim 1; SEQ ID NO 4550; 883pp + Sequence Listing; English.
XX
CC The invention relates to an isolated polynucleotide (ABN27253-ABN33262)
CC with cytostatic activity. The polynucleotide is used to produce a
CC polypeptide, to detect differentially expressed genes correlated with a
CC cancerous state of a mammalian cell and to inhibit tumour growth. The
CC polynucleotide is used as a probe in mapping and tissue profiling. The
CC encoded polypeptide and antibodies to the polypeptide can also be used
CC for therapeutic and diagnostic purposes. The polynucleotide is useful for
CC gene therapy.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 615 BP; 181 A; 123 C; 149 G; 162 T; 0 other;

Query Match 4.7%; Score 19; DB 24; Length 615;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 346 TTTCTTTTGAACACAAGTA 364
||||||||||||||||||
Db 205 TTTCTTTTGAACACAAGTA 223

RESULT 3
AAC46559
ID AAC46559 standard; DNA; 1444 BP.
XX
AC AAC46559;
XX
DT 18-OCT-2000 (first entry)
XX
DE Zea mays DNA fragment SEQ ID NO: 50583.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic;
KW pathway; promoter; termination sequence; corn; ss.
XX
OS Zea mays subsp. mays.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 4.7%; Score 19; DB 21; Length 1444;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    122 TTAAGCTTCGATGAGAAGA 140
          ||||||||||||||||||
Db   1121 TTAAGCTTCGATGAGAAGA 1139


RESULT 4
AAS69866/c
ID   AAS69866 standard; cDNA; 2016 BP.
XX
AC   AAS69866;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #5670.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG05679.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 5670; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2016 BP; 338 A; 631 C; 608 G; 439 T; 0 other;

  Query Match             4.7%;  Score 19;  DB 23;  Length 2016;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     39 AGACATGTGCCAGGAGAAG 57
          ||||||||||||||||||
Db   2004 AGACATGTGCCAGGAGAAG 1986


RESULT 5
AAS82153
ID   AAS82153 standard; cDNA; 2020 BP.
XX
AC   AAS82153;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #17957.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG17966.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 17957; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2020 BP; 439 A; 610 C; 634 G; 337 T; 0 other;

  Query Match             4.7%;  Score 19;  DB 23;  Length 2020;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     39 AGACATGTGCCAGGAGAAG 57
          ||||||||||||||||||
Db     13 AGACATGTGCCAGGAGAAG 31
```

```
RESULT 6
AAV89959
ID   AAV89959 standard; cDNA; 360 BP.
XX
AC   AAV89959;
XX
DT   15-FEB-1999  (first entry)
XX
DE   EST clone CS765.
XX
KW   Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
KW   tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;
KW   receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
KW   gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO9845436-A2.
XX
PD   15-OCT-1998.
XX
PF   10-APR-1998;   98WO-US06955.
XX
PR   10-APR-1997;   97US-0838821.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Jacobs K,  Lavallie ER,  McCoy JM,  Merberg D;
PI   Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1999-070077/06.
XX
PT   New polynucleotides encoding human secreted proteins - derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries.
XX
PS   Claim 1; Page 386; 618pp; English.
XX
CC   The present sequence represents a human expressed sequence tag (EST).
CC   The polynucleotide, which is a secreted EST, and the encoded protein
CC   are predicted to have useful biological activities which would make
CC   them suitable for treating, preventing or ameliorating medical
CC   conditions in humans and animals, although no supporting data is
CC   given. Suggested activities include nutritional activity, immune
CC   stimulating or suppressing activity, haematopoiesis regulating
CC   activity, tissue growth activity, activin/inhibin activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC   activity, receptor/ligand activity, anti-inflammatory activity,
CC   cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The polynucleotide may also be useful for gene therapy.
XX
SQ   Sequence 360 BP; 124 A; 79 C; 95 G; 62 T; 0 other;

  Query Match             4.5%;  Score 18;  DB 20;  Length 360;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     134 GAGAAGAAGGAAAGGAGA 151
           ||||||||||||||||||
Db     283 GAGAAGAAGGAAAGGAGA 300


RESULT 7
AAC99130
ID   AAC99130 standard; cDNA; 872 BP.
XX
AC   AAC99130;
XX
DT   09-MAR-2001  (first entry)
XX
DE   Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:358.
XX
KW   Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW   detection; diagnosis; identification; cytostatic; neuroprotective;
KW   nootropic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW   antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW   linkage analysis; tissue identification; tissue typing; forensic;
KW   neural; immune system; muscular; reproductive; gastrointestinal;
KW   pulmonary; cardiovascular; renal; proliferative; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055320-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05989.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-579444/54.
DR   P-PSDB; AAB54365.
XX
PT   New nucleic acid that is a pancreatic cancer antigen for preventing,
PT   treating, or ameliorating a medical condition, particular pancreatic
PT   cancer, or for use in assays for diagnosing a pathological condition -
XX
PS   Claim 1; Page 780; 1379pp; English.
XX
CC   AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC   proteins, called pancreatic cancer antigens, given in AAB54008 to
CC   AAB54466. The human pancreatic cancer antigens have cytostatic,
CC   neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,
CC   gynaecological, cardiant and antiinflammatory activities, and can be used
CC   in gene therapy. The polynucleotide and proteins can be used for
CC   preventing, treating, or ameliorating a medical condition or in assays
CC   for diagnosing a pathological condition or a susceptibility to one in a
CC   subject. Binding partners to the proteins and the activity of the
CC   proteins can be identified. The pancreatic cancer antigens can be used to
CC   detect, treat or prevent pancreatic disorders, especially cancer.
CC   Agonists and antagonists to the antigens can be screened for. The
CC   pancreatic cancer antigen polynucleotides can be used to design nucleic
CC   acid hybridisation probes that can be used in chromosome mapping, linkage
CC   analysis, tissue identification and/or typing and a variety of forensic
CC   and diagnostic methods. The proteins can be used to generate antibodies
CC   which are used to purify, detect and target the polypeptides, including
CC   both in vivo and in vitro diagnostic and therapeutic methods. The
CC   proteins can be used to treat or prevent neural, immune system, muscular,
CC   reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC   proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence 872 BP; 253 A; 186 C; 244 G; 183 T; 6 other;

  Query Match             4.5%;  Score 18;  DB 21;  Length 872;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     134 GAGAAGAAGGAAAGGAGA 151
           ||||||||||||||||||
Db     468 GAGAAGAAGGAAAGGAGA 485


RESULT 8
AAH06889
ID   AAH06889 standard; cDNA; 899 BP.
XX
AC   AAH06889;
XX
DT   26-JUN-2001  (first entry)
```

```
XX
DE   Human cDNA clone (5'-primer) SEQ ID NO:3724.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 1; SEQ ID 3724; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 899 BP; 246 A; 205 C; 265 G; 179 T; 4 other;

  Query Match             4.5%;  Score 18;  DB 22;  Length 899;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     134 GAGAAGAAGGAAAGGAGA 151
           ||||||||||||||||||
Db     742 GAGAAGAAGGAAAGGAGA 759


RESULT 9
AAA51824
ID   AAA51824 standard; cDNA; 1149 BP.
XX
AC   AAA51824;
XX
DT   09-JAN-2001  (first entry)
XX
DE   Human melanocortin 1 receptor protein MC-R1ESTc11 cDNA.
XX
KW   MC-R1; melanocortin 1 receptor; rhodopsin; G-protein coupled receptor;
KW   splice variant; MC-R1B; C-terminal extension; agonist; antagonist;
KW   modulator; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1149
FT                   /*tag= a
FT                   /product= Melanocortin_1_receptor_B_splice_variant
XX
PN   WO200039147-A1.
XX
PD   06-JUL-2000.
XX
PF   16-DEC-1999;   99WO-US29963.
XX
PR   23-DEC-1998;   98US-0113401.
XX
PA   (MERI ) MERCK & CO INC.
XX
PI   Howard AD, MacNeil DJ, Van Der Ploeg LHT;
XX
DR   WPI; 2000-452365/39.
DR   P-PSDB; AAY97019.
XX
PT   New nucleic acid encoding a human melanocortin 1 receptor protein
PT   (MC-R1B) for determining whether a substance is capable of binding to
PT   or activating human MC-R1B and identifying a substance that modulates
PT   MC-R1B receptor activity
XX
PS   Claim 2; Page 73; 101pp; English.
XX
CC   The invention concerns novel splice variants of the melanocortin 1
CC   receptor (MC-R1) protein belonging to the rhodopsin sub-family of
CC   G-protein coupled receptors. The splice variants, referred to as MC-R1B
CC   proteins, contain an intracellular domain with an additional 65 amino
CC   acid residues in comparison to previously disclosed human MC-R1, referred
CC   to as MC-R1A. Additionally, residue 317 of the MC-R1B proteins is Cys,
CC   whereas the C-terminal amino acid residue 317 of known MC-R1A proteins is
CC   Trp. The novel sequences can be used to determine whether a substance
CC   modulates MC-R1B receptor activity and to identify potential agonists or
CC   antagonists of MC-R1B. Pharmaceutical compositions comprising
CC   modulators of MC-R1B are used to treat or diagnose disorders involving
CC   inappropriate melanocortin expression or activity.
XX
SQ   Sequence 1149 BP; 188 A; 402 C; 319 G; 240 T; 0 other;

  Query Match             4.5%;  Score 18;  DB 21;  Length 1149;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     48 CCAGGAGAAGGGGCCTTG 65
          ||||||||||||||||||
Db   1032 CCAGGAGAAGGGGCCTTG 1049


RESULT 10
AAH14264/c
ID   AAH14264 standard; cDNA; 1205 BP.
XX
AC   AAH14264;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:11581.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
```

OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 11581; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 1205 BP; 313 A; 283 C; 309 G; 300 T; 0 other;

Query Match 4.5%; Score 18; DB 22; Length 1205;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 CCAGGAGAAGGGGCCTTG 65
||||||||||||||||||
Db 599 CCAGGAGAAGGGGCCTTG 582

RESULT 11
AAC69571/c
ID AAC69571 standard; DNA; 1984 BP.
XX
AC AAC69571;
XX
DT 31-JAN-2001 (first entry)
XX
DE Human secreted protein gene 60 clone HMSGK61.
XX
KW Immunosuppressive; antiarthritic; antirheumatic; antiproliferative;
KW cytostatic; cardiant; vasotropic; cerebroprotective; neuroprotective;
KW nootropic; antibacterial; virucide; fungicide; opthalmalogical;
KW vulnerary; gene therapy; infection; human; secreted protein; ss.
XX
OS Homo sapiens.
XX
PN WO200061623-A1.
XX
PD 19-OCT-2000.
XX
PF 06-APR-2000; 2000WO-US08979.
XX
PR 09-APR-1999; 99US-0128693.
PR 26-APR-1999; 99US-0130991.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Ni J, Komatsoulis GA, Rosen CA, Soppet DR, Shi Y;
PI Lafleur DW, Olsen HS, Ebner R, Florence KA, Moore PA, Birse CE;
PI Young PE;
XX
DR WPI; 2000-647418/62.
XX
PT New nucleic acid molecules encoding 62 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives -
XX
PS Claim 1; Page 574-575; 716pp; English.
XX
CC The invention relates to the isolation of genes (AAC69512-C69587)
CC encoding 62 human secreted proteins (AAB38321-B38396). The genes can be
CC used to generate fusion proteins by linking to the gene for the human
CC immunoglobulin G Fc portion (AAC69503) for increasing the stability of
CC the fusion protein as compared to the human protein only. The genes and
CC proteins are useful for preventing, ameliorating or treating medical
CC conditions, e.g. by protein or gene therapy. The genes are isolated
CC from a range of human tissues disclosed in the specification. The
CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) autoimmune diseases e.g.
CC rheumatoid arthritis; (b) hyperproliferative disorders e.g. neoplasms
CC of the breast or liver; (c) cardiovascular disorders e.g. cardiac
CC arrest; (d) cerebrovascular disorders e.g. cerebral ischemia; (e)
CC angiogenesis; (f) nervous system disorders e.g. Alzheimer's disease; (g)
CC infections caused by bacteria, viruses and fungi; and (h) ocular
CC disorders e.g. corneal infection. The polypeptides can also be used to
CC aid wound healing and epithelial cell proliferation, to prevent skin
CC aging due to sunburn, to maintain organs before transplantation, for
CC supporting cell culture of primary tissues, to regenerate tissues and in
CC chemotaxis.
XX
SQ Sequence 1984 BP; 422 A; 512 C; 440 G; 610 T; 0 other;

Query Match 4.5%; Score 18; DB 21; Length 1984;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 134 GAGAAGAAGGAAAGGAGA 151
||||||||||||||||||
Db 1090 GAGAAGAAGGAAAGGAGA 1073

RESULT 12
AAS73815
ID AAS73815 standard; cDNA; 3099 BP.
XX
AC AAS73815;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #9619.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

```
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG09628.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 9619; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3099 BP; 881 A; 794 C; 818 G; 606 T; 0 other;

  Query Match             4.5%;  Score 18;  DB 23;  Length 3099;
  Best Local Similarity   100.0%;  Pred. No. 47;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     134 GAGAAGAAGGAAAGGAGA 151
           ||||||||||||||||||
Db     724 GAGAAGAAGGAAAGGAGA 741


RESULT 13
AAS84712
ID   AAS84712 standard; cDNA; 3099 BP.
XX
AC   AAS84712;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #20516.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
```

```
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG20525.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 20516; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3099 BP; 881 A; 794 C; 818 G; 606 T; 0 other;

  Query Match             4.5%;  Score 18;  DB 23;  Length 3099;
  Best Local Similarity   100.0%;  Pred. No. 47;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     134 GAGAAGAAGGAAAGGAGA 151
           ||||||||||||||||||
Db     724 GAGAAGAAGGAAAGGAGA 741


RESULT 14
AAS72090
ID   AAS72090 standard; cDNA; 3137 BP.
XX
AC   AAS72090;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #7894.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
```

```
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG07903.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 7894; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3137 BP; 886 A; 791 C; 840 G; 620 T; 0 other;

  Query Match             4.5%;  Score 18;  DB 23;  Length 3137;
  Best Local Similarity   100.0%;  Pred. No. 47;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    134 GAGAAGAAGGAAAGGAGA 151
          ||||||||||||||||||
Db    724 GAGAAGAAGGAAAGGAGA 741


RESULT 15
AAH14592
ID   AAH14592 standard; cDNA; 3288 BP.
XX
AC   AAH14592;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:12199.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 12199; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 3288 BP; 945 A; 768 C; 882 G; 693 T; 0 other;

  Query Match             4.5%;  Score 18;  DB 22;  Length 3288;
  Best Local Similarity   100.0%;  Pred. No. 47;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    134 GAGAAGAAGGAAAGGAGA 151
          ||||||||||||||||||
Db    742 GAGAAGAAGGAAAGGAGA 759


Search completed: January  2, 2003, 16:07:17
Job time : 363 secs
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: January 2, 2003, 14:25:11 ; Search time 75 Seconds
(without alignments)
1643.787 Million cell updates/sec

Title: US-09-850-147-1
Perfect score: 402
Sequence: 1 ccgggccgggctcaccagct.....ggcataacttcctgtgttcc 402

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	17	4.2	489	4	US-09-134-001C-66	Sequence 66, Appl
	2	17	4.2	597	4	US-09-134-001C-1790	Sequence 1790, Ap
c	3	17	4.2	1393	4	US-09-737-698B-11	Sequence 11, Appl
	4	17	4.2	1766	5	PCT-US92-06532-2	Sequence 2, Appli
c	5	17	4.2	1946	4	US-09-737-698B-27	Sequence 27, Appl
	6	16	4.0	40	2	US-08-857-946-146	Sequence 146, App
	7	16	4.0	40	3	US-08-970-740-146	Sequence 146, App
c	8	16	4.0	99	1	US-07-918-953-3	Sequence 3, Appli
c	9	16	4.0	99	1	US-08-081-661-3	Sequence 3, Appli
	10	16	4.0	242	1	US-08-487-001A-32	Sequence 32, Appl
	11	16	4.0	242	2	US-08-630-822A-32	Sequence 32, Appl
	12	16	4.0	242	2	US-09-005-069-32	Sequence 32, Appl
c	13	16	4.0	275	1	US-07-918-953-12	Sequence 12, Appl
c	14	16	4.0	275	1	US-07-918-953-14	Sequence 14, Appl
c	15	16	4.0	275	1	US-08-081-661-12	Sequence 12, Appl
c	16	16	4.0	275	1	US-08-081-661-14	Sequence 14, Appl
c	17	16	4.0	276	1	US-07-918-953-16	Sequence 16, Appl
c	18	16	4.0	276	1	US-08-081-661-16	Sequence 16, Appl
c	19	16	4.0	298	1	US-07-826-928A-28	Sequence 28, Appl
	20	16	4.0	320	4	US-08-994-962-14	Sequence 14, Appl
c	21	16	4.0	330	5	PCT-US95-08596-1	Sequence 1, Appli
c	22	16	4.0	359	3	US-08-589-028-3	Sequence 3, Appli
c	23	16	4.0	359	3	US-08-784-582-3	Sequence 3, Appli
c	24	16	4.0	359	4	US-08-785-271-3	Sequence 3, Appli
c	25	16	4.0	416	4	US-08-945-140-5	Sequence 5, Appli
	26	16	4.0	457	4	US-08-994-962-12	Sequence 12, Appl
c	27	16	4.0	510	1	US-07-918-953-7	Sequence 7, Appli
c	28	16	4.0	510	1	US-08-081-661-7	Sequence 7, Appli
	29	16	4.0	514	4	US-08-994-962-11	Sequence 11, Appl
c	30	16	4.0	515	3	US-08-589-028-1	Sequence 1, Appli
c	31	16	4.0	515	3	US-08-784-582-1	Sequence 1, Appli
c	32	16	4.0	515	4	US-08-785-271-1	Sequence 1, Appli
c	33	16	4.0	598	4	US-09-537-696-11	Sequence 11, Appl
c	34	16	4.0	633	4	US-09-537-696-12	Sequence 12, Appl
c	35	16	4.0	666	4	US-09-537-696-13	Sequence 13, Appl
c	36	16	4.0	674	3	US-09-069-811-1	Sequence 1, Appli
c	37	16	4.0	674	4	US-09-543-106-1	Sequence 1, Appli
	38	16	4.0	860	4	US-08-994-962-1	Sequence 1, Appli
	39	16	4.0	1131	1	US-07-959-946-2	Sequence 2, Appli
	40	16	4.0	1131	1	US-08-333-577-2	Sequence 2, Appli
	41	16	4.0	1131	5	PCT-US92-08634-2	Sequence 2, Appli
c	42	16	4.0	1766	4	US-09-517-467B-18	Sequence 18, Appl
c	43	16	4.0	2492	3	US-08-695-191-3	Sequence 3, Appli
c	44	16	4.0	2492	3	US-08-682-080-3	Sequence 3, Appli
	45	16	4.0	2659	3	US-09-029-267-19	Sequence 19, Appl

ALIGNMENTS

RESULT 1
US-09-134-001C-66
; Sequence 66, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 66
; LENGTH: 489
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-66

Query Match 4.2%; Score 17; DB 4; Length 489;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 223 CTTTGACTTCAGGTGGA 239
||||||||||||||||
Db 332 CTTTGACTTCAGGTGGA 348

RESULT 2
US-09-134-001C-1790
; Sequence 1790, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1790
; LENGTH: 597
; TYPE: DNA

```
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1790

  Query Match             4.2%;  Score 17;  DB 4;  Length 597;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    250 AGGCAATATTTAAATCT 266
          |||||||||||||||||
Db    503 AGGCAATATTTAAATCT 519


RESULT 3
US-09-737-698B-11/c
; Sequence 11, Application US/09737698B
; Patent No. 6462258
; GENERAL INFORMATION:
;  APPLICANT: Fincher, Karen
;  APPLICANT:  Wilkinson, Jack
;  TITLE OF INVENTION: No. 6462258el Plant Expression Constructs
;  FILE REFERENCE: 38-21(51499)C
;  CURRENT APPLICATION NUMBER: US/09/737,698B
;  CURRENT FILING DATE:  2000-12-15
;  PRIOR APPLICATION NUMBER: US 60/171,173
;  PRIOR FILING DATE: 1999-12-16
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
;   LENGTH: 1393
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: promoter
;   LOCATION: (1)..(1393)
;   OTHER INFORMATION: Act11 promoter polynucleotide seqeunce and intron
US-09-737-698B-11

  Query Match             4.2%;  Score 17;  DB 4;  Length 1393;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    325 TCTGATGCATTCCATTT 341
          |||||||||||||||||
Db     30 TCTGATGCATTCCATTT 14


RESULT 4
PCT-US92-06532-2
; Sequence 2, Application PC/TUS9206532
;  GENERAL INFORMATION:
;    APPLICANT: Krause, James E.
;    TITLE OF INVENTION:  Human Substance P Receptor
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Scott J. Meyer, Monsanto Co., A3SD
;      STREET:  800 N. Lindbergh Blvd.
;      CITY:  St. Louis
;      STATE:  Missouri
;      COUNTRY:  U.S.A
;      ZIP:  63167
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/06532
;      FILING DATE:  19920805
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyer, Scott J.
;      REGISTRATION NUMBER:  25,275
;      REFERENCE/DOCKET NUMBER:  07-24(776)A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (314)694-3117
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1766 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  211..1431
PCT-US92-06532-2

  Query Match             4.2%;  Score 17;  DB 5;  Length 1766;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    329 ATGCATTCCATTTCTGG 345
          |||||||||||||||||
Db   1711 ATGCATTCCATTTCTGG 1727


RESULT 5
US-09-737-698B-27/c
; Sequence 27, Application US/09737698B
; Patent No. 6462258
; GENERAL INFORMATION:
;  APPLICANT: Fincher, Karen
;  APPLICANT:  Wilkinson, Jack
;  TITLE OF INVENTION: No. 6462258el Plant Expression Constructs
;  FILE REFERENCE: 38-21(51499)C
;  CURRENT APPLICATION NUMBER: US/09/737,698B
;  CURRENT FILING DATE:  2000-12-15
;  PRIOR APPLICATION NUMBER: US 60/171,173
;  PRIOR FILING DATE: 1999-12-16
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 27
;   LENGTH: 1946
;   TYPE: DNA
;   ORGANISM: artificial sequence
;   FEATURE:
;   NAME/KEY: promoter
;   LOCATION: (1)..(1946)
;   OTHER INFORMATION: chimeric promoter fusion FMV and Act11 polynucleotides +  Act1
;   OTHER INFORMATION: ntro
US-09-737-698B-27

  Query Match             4.2%;  Score 17;  DB 4;  Length 1946;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    325 TCTGATGCATTCCATTT 341
          |||||||||||||||||
Db    595 TCTGATGCATTCCATTT 579


RESULT 6
US-08-857-946-146
; Sequence 146, Application US/08857946
; Patent No. 5994075
;  GENERAL INFORMATION:
;    APPLICANT:  Goodfellow, P.N.
;    TITLE OF INVENTION:  METHODS FOR IDENTIFYING A MUTATION IN A
;    TITLE OF INVENTION:  GENE OF INTEREST
;    NUMBER OF SEQUENCES:  162
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Banner & Witcoff, Inc.
;      STREET:  75 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
```

```
;       ZIP:  02109-1807
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  WordPerfect 6.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/857,946
;       FILING DATE:  16-MAY-1997
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/60/017,824
;       FILING DATE:  17-MAY-1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Kathleen M. Williams
;       REGISTRATION NUMBER:  34,380
;       REFERENCE/DOCKET NUMBER:  3529/05573
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617-345-9100
;       TELEFAX:  617-345-9111
;   INFORMATION FOR SEQ ID NO:  146:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  40 bases
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  other nucleic acid
;     FEATURE:
;       NAME/KEY:  primer cKitMm300r
US-08-857-946-146

  Query Match             4.0%;  Score 16;  DB 2;  Length 40;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    214 CCATATTTCCTTTGAC 229
          ||||||||||||||||
Db     17 CCATATTTCCTTTGAC 32


RESULT 7
US-08-970-740-146
; Sequence 146, Application US/08970740
; Patent No. 6015670
;  GENERAL INFORMATION:
;    APPLICANT:  Goodfellow, P.N.
;    TITLE OF INVENTION:  METHODS FOR IDENTIFYING A MUTATION IN A
;    TITLE OF INVENTION:  GENE OF INTEREST
;    NUMBER OF SEQUENCES:  162
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Banner & Witcoff, Inc.
;      STREET:  28 State Street, 28th Floor
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect 6.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/970,740
;      FILING DATE:  14-NOV-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/857,946
;      FILING DATE:  16-MAY-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/017,824
;      FILING DATE:  17-MAY-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kathleen M. Williams
;      REGISTRATION NUMBER:  34,380
;      REFERENCE/DOCKET NUMBER:  3529/59829
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-227-7111
;      TELEFAX:  617-227-4399
;  INFORMATION FOR SEQ ID NO:  146:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  40 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;    FEATURE:
;      NAME/KEY:  primer cKitMm300r
US-08-970-740-146

  Query Match             4.0%;  Score 16;  DB 3;  Length 40;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    214 CCATATTTCCTTTGAC 229
          ||||||||||||||||
Db     17 CCATATTTCCTTTGAC 32


RESULT 8
US-07-918-953-3/c
; Sequence 3, Application US/07918953
; Patent No. 5268453
;  GENERAL INFORMATION:
;    APPLICANT:  ANDY, ROBIN J.
;    APPLICANT:  LARSON, ERIC R.
;    TITLE OF INVENTION:  TISSUE-SELECTIVE INSULIN ANALOGS
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 PAGE MILL ROAD
;      CITY:  PALO ALTO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/918,953
;      FILING DATE:  19920730
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BOZICEVIC, KARL
;      REGISTRATION NUMBER:  28,807
;      REFERENCE/DOCKET NUMBER:  21900-2027420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-677-7000
;      TELEFAX:  (415) 494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  99 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-07-918-953-3

  Query Match             4.0%;  Score 16;  DB 1;  Length 99;
  Best Local Similarity   100.0%;  Pred. No. 38;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    191 TCCAGGGCCAAGGGCT 206
          ||||||||||||||||
Db     92 TCCAGGGCCAAGGGCT 77
```

```
RESULT 9
US-08-081-661-3/c
; Sequence 3, Application US/08081661
; Patent No. 5446020
;  GENERAL INFORMATION:
;    APPLICANT:  ANDY, ROBIN J.
;    APPLICANT:  LARSON, ERIC R.
;    TITLE OF INVENTION:  TISSUE-SELECTIVE INSULIN ANALOGS
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 PAGE MILL ROAD
;      CITY:  PALO ALTO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/081,661
;      FILING DATE:  23-JUN-1993
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/918,953
;      FILING DATE:  30-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BOZICEVIC, KARL
;      REGISTRATION NUMBER:  28,807
;      REFERENCE/DOCKET NUMBER:  21900-2027420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-677-7000
;      TELEFAX:  (415) 494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  99 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-081-661-3

  Query Match             4.0%;  Score 16;  DB 1;  Length 99;
  Best Local Similarity   100.0%;  Pred. No. 38;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     191 TCCAGGGCCAAGGGCT 206
           ||||||||||||||||
Db      92 TCCAGGGCCAAGGGCT 77


RESULT 10
US-08-487-001A-32
; Sequence 32, Application US/08487001A
; Patent No. 5795862
;  GENERAL INFORMATION:
;    APPLICANT:  FRANK, GLENN R.
;    APPLICANT:  HUNTER, SHIRLEY WU
;    APPLICANT:  WALLENFELS, LYNDA
;    TITLE OF INVENTION:  NOVEL ECTOPARASITE SALIVA
;    TITLE OF INVENTION:  PROTEINS AND APPARATUS TO COLLECT SUCH PROTEINS
;    NUMBER OF SEQUENCES:  54
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheridan Ross & McIntosh
;      STREET:  1700 Lincoln Street, Suite 3500
;      CITY:  Denver
;      STATE:  Colorado
;      COUNTRY:  U.S.A.
;      ZIP:  80203
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/487,001A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Verser, Carol Talkington
;      REGISTRATION NUMBER:  37,459
;      REFERENCE/DOCKET NUMBER:  2618-17-C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 863-9700
;      TELEFAX:  (303) 863-0223
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  242 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3..242
US-08-487-001A-32

  Query Match             4.0%;  Score 16;  DB 1;  Length 242;
  Best Local Similarity   100.0%;  Pred. No. 39;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      87 AATTATTGCTCAGATG 102
           ||||||||||||||||
Db     129 AATTATTGCTCAGATG 144


RESULT 11
US-08-630-822A-32
; Sequence 32, Application US/08630822A
; Patent No. 5840695
;  GENERAL INFORMATION:
;    APPLICANT:  FRANK, GLENN R.
;    APPLICANT:  HUNTER, SHIRLEY WU
;    APPLICANT:  WALLENFELS, LYNDA
;    TITLE OF INVENTION:  NOVEL ECTOPARASITE SALIVA PROTEINS
;    TITLE OF INVENTION:  AND APPARATUS TO COLLECT SUCH PROTEINS
;    NUMBER OF SEQUENCES:  107
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheridan Ross P.C.
;      STREET:  1700 Lincoln Street, Suite 3500
;      CITY:  Denver
;      STATE:  Colorado
;      COUNTRY:  U.S.A.
;      ZIP:  80203
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/630,822A
;      FILING DATE:  11-APR-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CONNELL, GARY J.
;      REGISTRATION NUMBER:  32,020
;      REFERENCE/DOCKET NUMBER:  2618-17-C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 863-9700
;      TELEFAX:  (303) 863-0223
;  INFORMATION FOR SEQ ID NO:  32:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  242 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3..242
US-08-630-822A-32

  Query Match             4.0%;  Score 16;  DB 2;  Length 242;
  Best Local Similarity   100.0%;  Pred. No. 39;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      87 AATTATTGCTCAGATG 102
           ||||||||||||||||
Db     129 AATTATTGCTCAGATG 144


RESULT 12
US-09-005-069-32
; Sequence 32, Application US/09005069
; Patent No. 5932470
;  GENERAL INFORMATION:
;    APPLICANT:  FRANK, GLENN R.
;    APPLICANT:  HUNTER, SHIRLEY WU
;    APPLICANT:  WALLENFELS, LYNDA
;    TITLE OF INVENTION:  NOVEL ECTOPARASITE SALIVA PROTEINS
;    TITLE OF INVENTION:  AND APPARATUS TO COLLECT SUCH PROTEINS
;    NUMBER OF SEQUENCES:  107
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheridan Ross P.C.
;      STREET:  1700 Lincoln Street, Suite 3500
;      CITY:  Denver
;      STATE:  Colorado
;      COUNTRY:  U.S.A.
;      ZIP:  80203
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/005,069
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/630,822
;      FILING DATE:  11-APR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CONNELL, GARY J.
;      REGISTRATION NUMBER:  32,020
;      REFERENCE/DOCKET NUMBER:  2618-17-C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 863-9700
;      TELEFAX:  (303) 863-0223
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  242 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3..242
US-09-005-069-32

  Query Match             4.0%;  Score 16;  DB 2;  Length 242;
  Best Local Similarity   100.0%;  Pred. No. 39;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      87 AATTATTGCTCAGATG 102
           ||||||||||||||||
Db     129 AATTATTGCTCAGATG 144


RESULT 13
US-07-918-953-12/c
; Sequence 12, Application US/07918953
; Patent No. 5268453
;  GENERAL INFORMATION:
;    APPLICANT:  ANDY, ROBIN J.
;    APPLICANT:  LARSON, ERIC R.
;    TITLE OF INVENTION:  TISSUE-SELECTIVE INSULIN ANALOGS
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 PAGE MILL ROAD
;      CITY:  PALO ALTO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/918,953
;      FILING DATE:  19920730
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BOZICEVIC, KARL
;      REGISTRATION NUMBER:  28,807
;      REFERENCE/DOCKET NUMBER:  21900-2027420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-677-7000
;      TELEFAX:  (415) 494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  275 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  209..271
;      OTHER INFORMATION:  /note= "The insulin analog
;      OTHER INFORMATION:  is numbered from amino acid number 1 of the
;      OTHER INFORMATION:  A chain."
US-07-918-953-12

  Query Match             4.0%;  Score 16;  DB 1;  Length 275;
  Best Local Similarity   100.0%;  Pred. No. 39;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     191 TCCAGGGCCAAGGGCT 206
           ||||||||||||||||
Db     189 TCCAGGGCCAAGGGCT 174


RESULT 14
US-07-918-953-14/c
; Sequence 14, Application US/07918953
; Patent No. 5268453
;  GENERAL INFORMATION:
;    APPLICANT:  ANDY, ROBIN J.
;    APPLICANT:  LARSON, ERIC R.
;    TITLE OF INVENTION:  TISSUE-SELECTIVE INSULIN ANALOGS
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
```

```
;       STREET:  755 PAGE MILL ROAD
;       CITY:  PALO ALTO
;       STATE:  CALIFORNIA
;       COUNTRY:  USA
;       ZIP:  94304-1018
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/918,953
;       FILING DATE:  19920730
;       CLASSIFICATION:  514
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BOZICEVIC, KARL
;       REGISTRATION NUMBER:  28,807
;       REFERENCE/DOCKET NUMBER:  21900-2027420
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415-677-7000
;       TELEFAX:  (415) 494-0792
;       TELEX:  706141
;   INFORMATION FOR SEQ ID NO:  14:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  275 base pairs
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  209..271
;       OTHER INFORMATION:  /note= "The insulin analog
;       OTHER INFORMATION:  is numbered from amino acid number 1 of the
;       OTHER INFORMATION:  A chain."
US-07-918-953-14

  Query Match             4.0%;  Score 16;  DB 1;  Length 275;
  Best Local Similarity   100.0%;  Pred. No. 39;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     191 TCCAGGGCCAAGGGCT 206
           ||||||||||||||||
Db     189 TCCAGGGCCAAGGGCT 174


RESULT 15
US-08-081-661-12/c
; Sequence 12, Application US/08081661
; Patent No. 5446020
;  GENERAL INFORMATION:
;    APPLICANT:  ANDY, ROBIN J.
;    APPLICANT:  LARSON, ERIC R.
;    TITLE OF INVENTION:  TISSUE-SELECTIVE INSULIN ANALOGS
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 PAGE MILL ROAD
;      CITY:  PALO ALTO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/081,661
;      FILING DATE:  23-JUN-1993
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/918,953
;      FILING DATE:  30-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BOZICEVIC, KARL
;      REGISTRATION NUMBER:  28,807
;      REFERENCE/DOCKET NUMBER:  21900-2027420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-677-7000
;      TELEFAX:  (415) 494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  275 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  209..271
;      OTHER INFORMATION:  /note= "The insulin analog
;      OTHER INFORMATION:  is numbered from amino acid number 1 of the
;      OTHER INFORMATION:  A chain."
US-08-081-661-12

  Query Match             4.0%;  Score 16;  DB 1;  Length 275;
  Best Local Similarity   100.0%;  Pred. No. 39;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     191 TCCAGGGCCAAGGGCT 206
           ||||||||||||||||
Db     189 TCCAGGGCCAAGGGCT 174


Search completed: January  2, 2003, 16:01:03
Job time : 78 secs
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: January 2, 2003, 16:01:07 ; Search time 84 Seconds
(without alignments)
2069.842 Million cell updates/sec

Title: US-09-850-147-1
Perfect score: 402
Sequence: 1 ccgggccgggctcaccagct.....ggcataacttcctgtgttcc 402

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 381593 seqs, 216252194 residues

Word size : 0

Total number of hits satisfying chosen parameters: 763186

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	22	5.5	574	12	US-10-062-254-335	Sequence 335, App
	2	18	4.5	872	10	US-09-925-297-358	Sequence 358, App
c	3	18	4.5	2000	9	US-09-938-842A-3855	Sequence 3855, Ap
	4	18	4.5	4623	10	US-09-864-864-288	Sequence 288, App
c	5	18	4.5	25277	10	US-09-764-878-285	Sequence 285, App
c	6	18	4.5	25277	10	US-09-764-860-1148	Sequence 1148, Ap
c	7	18	4.5	25279	10	US-09-764-878-284	Sequence 284, App
c	8	18	4.5	25279	10	US-09-764-860-1147	Sequence 1147, Ap
	9	17	4.2	179	10	US-09-954-456-1496	Sequence 1496, Ap
	10	17	4.2	179	10	US-09-880-107-2996	Sequence 2996, Ap
c	11	17	4.2	219	9	US-09-796-692-7632	Sequence 7632, Ap
c	12	17	4.2	260	10	US-09-878-574-11983	Sequence 11983, A
c	13	17	4.2	498	10	US-09-925-297-346	Sequence 346, App
c	14	17	4.2	1183	9	US-10-008-118A-33	Sequence 33, Appl
c	15	17	4.2	1183	10	US-09-443-704-33	Sequence 33, Appl
c	16	17	4.2	1393	10	US-09-737-626A-11	Sequence 11, Appl
c	17	17	4.2	1946	10	US-09-737-626A-27	Sequence 27, Appl
	18	17	4.2	2982	9	US-09-938-842A-398	Sequence 398, App
	19	17	4.2	3597	9	US-10-098-841-84	Sequence 84, Appl
	20	16	4.0	94	10	US-09-864-761-30732	Sequence 30732, A
c	21	16	4.0	108	10	US-09-563-817-655	Sequence 655, App
c	22	16	4.0	138	10	US-09-974-300-5722	Sequence 5722, Ap
	23	16	4.0	201	10	US-09-789-529-20	Sequence 20, Appl
	24	16	4.0	228	10	US-09-867-701-7968	Sequence 7968, Ap
	25	16	4.0	242	7	US-08-809-423A-32	Sequence 32, Appl
	26	16	4.0	269	10	US-09-878-574-7917	Sequence 7917, Ap
	27	16	4.0	320	10	US-09-899-917-14	Sequence 14, Appl
c	28	16	4.0	350	9	US-09-796-692-7386	Sequence 7386, Ap
c	29	16	4.0	351	10	US-09-280-030-49	Sequence 49, Appl
c	30	16	4.0	374	9	US-09-938-842A-4736	Sequence 4736, Ap
	31	16	4.0	385	10	US-09-864-761-14179	Sequence 14179, A
c	32	16	4.0	390	10	US-09-280-030-48	Sequence 48, Appl
c	33	16	4.0	395	10	US-09-960-352-10581	Sequence 10581, A
	34	16	4.0	420	10	US-09-864-761-10796	Sequence 10796, A
c	35	16	4.0	428	10	US-09-783-590-8704	Sequence 8704, Ap
c	36	16	4.0	450	9	US-09-804-409A-10	Sequence 10, Appl
	37	16	4.0	457	10	US-09-899-917-12	Sequence 12, Appl
c	38	16	4.0	468	9	US-09-796-692-4911	Sequence 4911, Ap
c	39	16	4.0	468	9	US-09-796-692-9012	Sequence 9012, Ap
c	40	16	4.0	484	10	US-09-867-550-1781	Sequence 1781, Ap
c	41	16	4.0	512	9	US-09-728-444-427	Sequence 427, App
	42	16	4.0	514	10	US-09-899-917-11	Sequence 11, Appl
c	43	16	4.0	531	10	US-09-864-761-6877	Sequence 6877, Ap
	44	16	4.0	576	10	US-09-864-761-9132	Sequence 9132, Ap
c	45	16	4.0	581	10	US-09-864-761-9876	Sequence 9876, Ap

ALIGNMENTS

RESULT 1
US-10-062-254-335
; Sequence 335, Application US/10062254
; Patent No. US20020138882A1
; GENERAL INFORMATION:
; APPLICANT: Cahoon, Edgar B
; APPLICANT: Cahoon, Rebecca E
; APPLICANT: Falco, Saverio Carl
; APPLICANT: Fang, Yiwen
; APPLICANT: Hantke, Sabine S.
; APPLICANT: Lee, Jian-Ming
; APPLICANT: Li, Zhongsen
; APPLICANT: Miao, Guo-Hua
; APPLICANT: Morgante, Michele
; APPLICANT: Niu, Xiping
; APPLICANT: Odell, Joan
; APPLICANT: Rafalski, Antoni
; APPLICANT: Sakai, Hajime
; APPLICANT: Zheng, Peizhong
; APPLICANT: Zhu, Qun
; TITLE OF INVENTION: Polynucleotides Encoding Proteins Involved In Plant Metabolism
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/062,254
; CURRENT FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 09/630,346
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 60/146511
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 60/156006
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 60/156899
; PRIOR FILING DATE: 1999-09-30
; PRIOR APPLICATION NUMBER: 60/157287
; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: 60/169767
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 60/171054
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: 60/172958
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/171515
; PRIOR FILING DATE: 1999-12-22

; PRIOR APPLICATION NUMBER: 60/173535
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 375
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 335
; LENGTH: 574
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (495)
; NAME/KEY: unsure
; LOCATION: (511)
; NAME/KEY: unsure
; LOCATION: (528)
; NAME/KEY: unsure
; LOCATION: (538)
US-10-062-254-335

Query Match 5.5%; Score 22; DB 12; Length 574;
Best Local Similarity 100.0%; Pred. No. 0.046;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 131 GATGAGAAGAAGGAAAGGAGAT 152
||||||||||||||||||||||
Db 95 GATGAGAAGAAGGAAAGGAGAT 116

RESULT 2
US-09-925-297-358
; Sequence 358, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 358
; LENGTH: 872
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (803)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (813)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (871)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-297-358

Query Match 4.5%; Score 18; DB 10; Length 872;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 134 GAGAAGAAGGAAAGGAGA 151
||||||||||||||||||
Db 468 GAGAAGAAGGAAAGGAGA 485

RESULT 3
US-09-938-842A-3855/c
; Sequence 3855, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIP1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 3855
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-3855

Query Match 4.5%; Score 18; DB 9; Length 2000;
Best Local Similarity 100.0%; Pred. No. 6.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 123 TAAGCTTCGATGAGAAGA 140
||||||||||||||||||
Db 1630 TAAGCTTCGATGAGAAGA 1613

RESULT 4
US-09-864-864-288
; Sequence 288, Application US/09864864
; Patent No. US20020102679A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Secrist, Heather
; APPLICANT: Lodes, Michael J.
; APPLICANT: Algate, Paul A.
; APPLICANT: Fling, Steve P.
; APPLICANT: Mannion, Jane
; APPLICANT: Benson, Darin R.
; APPLICANT: Carter, Darrick
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.523
; CURRENT APPLICATION NUMBER: US/09/864,864
; CURRENT FILING DATE: 2001-05-23
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 288
; LENGTH: 4623
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-864-864-288

Query Match 4.5%; Score 18; DB 10; Length 4623;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 134 GAGAAGAAGGAAAGGAGA 151
||||||||||||||||||
Db 1072 GAGAAGAAGGAAAGGAGA 1089

RESULT 5
US-09-764-878-285/c
; Sequence 285, Application US/09764878

```
; Patent No. US20020090615A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA121
;  CURRENT APPLICATION NUMBER: US/09/764,878
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 428
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 285
;   LENGTH: 25277
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-878-285

  Query Match             4.5%;  Score 18;  DB 10;  Length 25277;
  Best Local Similarity   100.0%;  Pred. No. 8.4;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    241 AATTTTAGAAGGCAATAT 258
          ||||||||||||||||||
Db   2821 AATTTTAGAAGGCAATAT 2804


RESULT 6
US-09-764-860-1148/c
; Sequence 1148, Application US/09764860
; Patent No. US20020094953A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC008
;  CURRENT APPLICATION NUMBER: US/09/764,860
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1198
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1148
;   LENGTH: 25277
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-860-1148

  Query Match             4.5%;  Score 18;  DB 10;  Length 25277;
  Best Local Similarity   100.0%;  Pred. No. 8.4;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    241 AATTTTAGAAGGCAATAT 258
          ||||||||||||||||||
Db   2821 AATTTTAGAAGGCAATAT 2804


RESULT 7
US-09-764-878-284/c
; Sequence 284, Application US/09764878
; Patent No. US20020090615A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA121
;  CURRENT APPLICATION NUMBER: US/09/764,878
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 428
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 284
;   LENGTH: 25279
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-878-284

  Query Match             4.5%;  Score 18;  DB 10;  Length 25279;
  Best Local Similarity   100.0%;  Pred. No. 8.4;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    241 AATTTTAGAAGGCAATAT 258
          ||||||||||||||||||
Db   2821 AATTTTAGAAGGCAATAT 2804


RESULT 8
US-09-764-860-1147/c
; Sequence 1147, Application US/09764860
; Patent No. US20020094953A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC008
;  CURRENT APPLICATION NUMBER: US/09/764,860
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1198
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1147
;   LENGTH: 25279
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-860-1147

  Query Match             4.5%;  Score 18;  DB 10;  Length 25279;
  Best Local Similarity   100.0%;  Pred. No. 8.4;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    241 AATTTTAGAAGGCAATAT 258
          ||||||||||||||||||
Db   2821 AATTTTAGAAGGCAATAT 2804


RESULT 9
US-09-954-456-1496
; Sequence 1496, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
;  APPLICANT: Young, Paul
;  TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
;  TITLE OF INVENTION:  Sets
;  FILE REFERENCE: 689290-76
;  CURRENT APPLICATION NUMBER: US/09/954,456
;  CURRENT FILING DATE:  2001-09-18
;  PRIOR APPLICATION NUMBER: US/60/233,617
;  PRIOR FILING DATE: 2000-09-18
;  PRIOR APPLICATION NUMBER: US/60/234,052
;  PRIOR FILING DATE: 2000-09-20
;  PRIOR APPLICATION NUMBER: US/60/234,923
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,134
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,637
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,638
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,711
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,720
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,840
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,863
;  PRIOR FILING DATE: 2000-09-27
;  NUMBER OF SEQ ID NOS: 2276
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1496
;   LENGTH: 179
;   TYPE: DNA
;   ORGANISM: Homo sapiens
```

```
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: n=a,t,g or c
US-09-954-456-1496

  Query Match             4.2%;  Score 17;  DB 10;  Length 179;
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    135 AGAAGAAGGAAAGGAGA 151
          |||||||||||||||||
Db     71 AGAAGAAGGAAAGGAGA 87


RESULT 10
US-09-880-107-2996
; Sequence 2996, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
;  APPLICANT: Horne, Darci T.
;  APPLICANT:  Vockley, Joseph G.
;  APPLICANT:  Scherf, Uwe
;  APPLICANT:  Gene Logic, Inc.
;  TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
;  FILE REFERENCE: 44921-5028-WO
;  CURRENT APPLICATION NUMBER: US/09/880,107
;  CURRENT FILING DATE:  2001-06-14
;  PRIOR APPLICATION NUMBER: US 60/211,379
;  PRIOR FILING DATE: 2000-06-14
;  PRIOR APPLICATION NUMBER: US 60/237,054
;  PRIOR FILING DATE: 2000-10-02
;  NUMBER OF SEQ ID NOS: 3950
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2996
;   LENGTH: 179
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20020142981A1 R95966
;   NAME/KEY: unsure
;   LOCATION: (1)..(179)
;   OTHER INFORMATION: n = a or c or g or t
US-09-880-107-2996

  Query Match             4.2%;  Score 17;  DB 10;  Length 179;
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    135 AGAAGAAGGAAAGGAGA 151
          |||||||||||||||||
Db     71 AGAAGAAGGAAAGGAGA 87


RESULT 11
US-09-796-692-7632/c
; Sequence 7632, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
;  APPLICANT: Gaiger, Alexander
;  APPLICANT:  Algate, Paul A.
;  APPLICANT:  Mannion, Jane
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
;  TITLE OF INVENTION:  HEMATOLOGICAL MALIGNANCIES
;  FILE REFERENCE: 2077.001200
;  CURRENT APPLICATION NUMBER: US/09/796,692
;  CURRENT FILING DATE:  2001-03-01
;  PRIOR APPLICATION NUMBER: 60/186,126
;  PRIOR FILING DATE: 2000-03-01
;  PRIOR APPLICATION NUMBER: 60/190,479
;  PRIOR FILING DATE: 2000-03-17
;  PRIOR APPLICATION NUMBER: 60/200,545
;  PRIOR FILING DATE: 2000-04-27
;  PRIOR APPLICATION NUMBER: 60/200,303
;  PRIOR FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: 60/200,779
;  PRIOR FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: 60/200,999
;  PRIOR FILING DATE: 2000-05-01
;  PRIOR APPLICATION NUMBER: 60/202,084
;  PRIOR FILING DATE: 2000-05-04
;  PRIOR APPLICATION NUMBER: 60/206,201
;  PRIOR FILING DATE: 2000-05-22
;  PRIOR APPLICATION NUMBER: 60/218,950
;  PRIOR FILING DATE: 2000-07-14
;  PRIOR APPLICATION NUMBER: 60/222,903
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: 60/223,416
;  PRIOR FILING DATE: 2000-08-04
;  PRIOR APPLICATION NUMBER: 60/223,378
;  PRIOR FILING DATE: 2000-08-07
;  NUMBER OF SEQ ID NOS: 9597
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7632
;   LENGTH: 219
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (151)
;   OTHER INFORMATION: n=A,T,C or G
;   NAME/KEY: unsure
;   LOCATION: (187)
;   OTHER INFORMATION: n=A,T,C or G
;   NAME/KEY: unsure
;   LOCATION: (211)
;   OTHER INFORMATION: n=A,T,C or G
;   NAME/KEY: unsure
;   LOCATION: (216)
;   OTHER INFORMATION: n=A,T,C or G
US-09-796-692-7632

  Query Match             4.2%;  Score 17;  DB 9;  Length 219;
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    136 GAAGAAGGAAAGGAGAT 152
          |||||||||||||||||
Db    176 GAAGAAGGAAAGGAGAT 160


RESULT 12
US-09-878-574-11983/c
; Sequence 11983, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
;  APPLICANT: Byrum, Joseph R.
;  APPLICANT:  La Rosa, Thomas J.
;  APPLICANT:  Thompson, Michael D.
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(15401)B
;  CURRENT APPLICATION NUMBER: US/09/878,574
;  CURRENT FILING DATE:  2001-12-21
;  PRIOR APPLICATION NUMBER: 09/333,535
;  PRIOR FILING DATE: 1999-06-14
;  NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 11983
;   LENGTH: 260
;   TYPE: DNA
;   ORGANISM: Glycine max
;   OTHER INFORMATION: Clone ID: 701065166H1
US-09-878-574-11983

  Query Match             4.2%;  Score 17;  DB 10;  Length 260;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy    353 TGAACACAAGTACAAAA 369
          |||||||||||||||||
Db     37 TGAACACAAGTACAAAA 21


RESULT 13
US-09-925-297-346/c
; Sequence 346, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA105
;  CURRENT APPLICATION NUMBER: US/09/925,297
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05989
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 928
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 346
;   LENGTH: 498
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc feature
;   LOCATION: (252)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (493)
;   OTHER INFORMATION: n equals a,t,g, or c.
US-09-925-297-346

  Query Match             4.2%;  Score 17;  DB 10;  Length 498;
  Best Local Similarity   100.0%;  Pred. No. 21;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    190 TTCCAGGGCCAAGGGCT 206
          |||||||||||||||||
Db    300 TTCCAGGGCCAAGGGCT 284


RESULT 14
US-10-008-118A-33/c
; Sequence 33, Application US/10008118A
; Publication No. US20020187539A1
; GENERAL INFORMATION:
;  APPLICANT: Cahoon, Rebecca E.
;  APPLICANT:  Weng, Zude
;  TITLE OF INVENTION: Plant MYB-Related Transcription Factors
;  FILE REFERENCE: BB1280 USDIV
;  CURRENT APPLICATION NUMBER: US/10/008,118A
;  CURRENT FILING DATE:  2001-12-05
;  PRIOR APPLICATION NUMBER: 60/109,294
;  PRIOR FILING DATE: 1998-11-20
;  NUMBER OF SEQ ID NOS: 50
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 33
;   LENGTH: 1183
;   TYPE: DNA
;   ORGANISM: Triticum aestivum
US-10-008-118A-33

  Query Match             4.2%;  Score 17;  DB 9;  Length 1183;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    137 AAGAAGGAAAGGAGATG 153
          |||||||||||||||||
Db   1089 AAGAAGGAAAGGAGATG 1073


RESULT 15
US-09-443-704-33/c
; Sequence 33, Application US/09443704
; Patent No. US20020066120A1
; GENERAL INFORMATION:
;  APPLICANT: Cahoon, Rebecca E.
;  APPLICANT: Liu, Zhan-Bin
;  APPLICANT: Odell, Joan
;  APPLICANT: Rafalski, Antoni
;  APPLICANT: Shi, June
;  APPLICANT: Weng, Zude
;  TITLE OF INVENTION: Plant MYB-Related Transcription Factors
;  FILE REFERENCE: BB1280 US NA
;  CURRENT APPLICATION NUMBER: US/09/443,704
;  CURRENT FILING DATE: 1999-11-19
;  EARLIER APPLICATION NUMBER: 60/109,294
;  EARLIER FILING DATE: No. US20020066120A1ember 20, 1998
;  NUMBER OF SEQ ID NOS: 50
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 33
;   LENGTH: 1183
;   TYPE: DNA
;   ORGANISM: Triticum aestivum
US-09-443-704-33

  Query Match             4.2%;  Score 17;  DB 10;  Length 1183;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    137 AAGAAGGAAAGGAGATG 153
          |||||||||||||||||
Db   1089 AAGAAGGAAAGGAGATG 1073


Search completed: January  2, 2003, 17:49:29
Job time : 134 secs
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: January 2, 2003, 16:07:27 ; Search time 147 Seconds
(without alignments)
2271.753 Million cell updates/sec

Title: US-09-850-147-1
Perfect score: 402
Sequence: 1 ccgggccgggctcaccagct.....ggcataacttcctgtgttcc 402

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 655294 seqs, 415357811 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1310588

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	18	4.5	19	6	US-10-310-188-10261	Sequence 10261, A
c 2	18	4.5	319608	5	US-09-539-333D-1	GENERAL INFORMATI
c 3	18	4.5	319608	5	US-09-539-333D-1	Sequence 1, Appli
4	17	4.2	236	5	US-09-534-850-21604	Sequence 21604, A
5	17	4.2	238	5	US-09-534-850-20438	Sequence 20438, A
c 6	17	4.2	254	5	US-09-513-999C-20190	Sequence 20190, A
c 7	17	4.2	261	5	US-09-513-999C-14719	Sequence 14719, A
8	17	4.2	324	5	US-09-513-999C-33198	Sequence 33198, A
9	17	4.2	489	6	US-10-092-411A-66	Sequence 66, Appl
10	17	4.2	597	6	US-10-092-411A-1790	Sequence 1790, Ap
c 11	17	4.2	1165	5	US-09-724-676-25002	Sequence 25002, A
c 12	17	4.2	1165	5	US-09-724-676A-25002	Sequence 25002, A
c 13	17	4.2	1199	5	US-09-724-676-25006	Sequence 25006, A
c 14	17	4.2	1199	5	US-09-724-676A-25006	Sequence 25006, A
c 15	17	4.2	1203	5	US-09-724-676-25004	Sequence 25004, A
c 16	17	4.2	1203	5	US-09-724-676A-25004	Sequence 25004, A
c 17	17	4.2	1237	5	US-09-724-676-25008	Sequence 25008, A
c 18	17	4.2	1237	5	US-09-724-676A-25008	Sequence 25008, A
c 19	17	4.2	1285	5	US-09-724-676-25003	Sequence 25003, A
c 20	17	4.2	1285	5	US-09-724-676A-25003	Sequence 25003, A
c 21	17	4.2	1319	5	US-09-724-676-25007	Sequence 25007, A
c 22	17	4.2	1319	5	US-09-724-676A-25007	Sequence 25007, A
c 23	17	4.2	1355	5	US-09-724-676-25005	Sequence 25005, A
c 24	17	4.2	1355	5	US-09-724-676A-25005	Sequence 25005, A
c 25	17	4.2	1389	5	US-09-724-676-25009	Sequence 25009, A
c 26	17	4.2	1389	5	US-09-724-676A-25009	Sequence 25009, A
c 27	17	4.2	1536	5	US-09-724-676-24992	Sequence 24992, A
c 28	17	4.2	1536	5	US-09-724-676A-24992	Sequence 24992, A
c 29	17	4.2	1574	5	US-09-724-676-24994	Sequence 24994, A
c 30	17	4.2	1574	5	US-09-724-676A-24994	Sequence 24994, A
c 31	17	4.2	1595	5	US-09-724-676-24987	Sequence 24987, A
c 32	17	4.2	1595	5	US-09-724-676A-24987	Sequence 24987, A
c 33	17	4.2	1633	5	US-09-724-676-24989	Sequence 24989, A
c 34	17	4.2	1633	5	US-09-724-676A-24989	Sequence 24989, A
c 35	17	4.2	1656	5	US-09-724-676-24993	Sequence 24993, A
c 36	17	4.2	1656	5	US-09-724-676A-24993	Sequence 24993, A
c 37	17	4.2	1715	5	US-09-724-676-24988	Sequence 24988, A
c 38	17	4.2	1715	5	US-09-724-676A-24988	Sequence 24988, A
c 39	17	4.2	1726	5	US-09-724-676-24995	Sequence 24995, A
c 40	17	4.2	1726	5	US-09-724-676A-24995	Sequence 24995, A
c 41	17	4.2	1785	5	US-09-724-676-24990	Sequence 24990, A
c 42	17	4.2	1785	5	US-09-724-676A-24990	Sequence 24990, A
43	17	4.2	1986	5	US-09-724-676-23591	Sequence 23591, A
44	17	4.2	1986	5	US-09-724-676A-23591	Sequence 23591, A
45	17	4.2	263744	6	US-10-229-834A-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-10-310-188-10261/c
; Sequence 10261, Application US/10310188
; GENERAL INFORMATION:
; APPLICANT: RosettaGemonics
; TITLE OF INVENTION: BIOINFORMATICALLY DETECTABLE GROUP OF NOVEL VIRAL REGULATORY G
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 47487
; CURRENT APPLICATION NUMBER: US/10/310,188
; CURRENT FILING DATE: 2002-12-19
; NUMBER OF SEQ ID NOS: 86841
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10261
; LENGTH: 19
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-310-188-10261

Query Match 4.5%; Score 18; DB 6; Length 19;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 134 GAGAAGAAGGAAAGGAGA 151
||||||||||||||||||
Db 19 GAGAAGAAGGAAAGGAGA 2

RESULT 2
US-09-539-333D-1/c
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Bihain, Bernard
; APPLICANT: Essioux, Laurent
; TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
; FILE REFERENCE: GENSET.047AUS
; CURRENT APPLICATION NUMBER: US/09/539,333D
; CURRENT FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US 60/126,903
; PRIOR FILING DATE: 1999-03-30
; PRIOR APPLICATION NUMBER: US 60/131,971
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/132,065
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/143,928
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: US 60/145,915

; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: US 60/146,453
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/146,452
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/162,288
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: US 09/416,384
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: Patent.pm
; FEATURE:
; NAME/KEY: allele
; LOCATION: 110222
; OTHER INFORMATION: polymorphic base G or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 111978
; OTHER INFORMATION: polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 112468
; OTHER INFORMATION: polymorphic base G or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 117324..117327
; OTHER INFORMATION: deletion ACTT
; FEATURE:
; NAME/KEY: allele
; LOCATION: 118972
; OTHER INFORMATION: polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 119160..119161
; OTHER INFORMATION: deletion TT
; FEATURE:
; NAME/KEY: allele
; LOCATION: 119316
; OTHER INFORMATION: polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 119321
; OTHER INFORMATION: polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 119526
; OTHER INFORMATION: polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 120573
; OTHER INFORMATION: polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 121527
; OTHER INFORMATION: polymorphic base A or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 126105
; OTHER INFORMATION: polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 129789
; OTHER INFORMATION: polymorphic base C or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 130777
; OTHER INFORMATION: polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 136942..136944
; OTHER INFORMATION: deletion ATT
; FEATURE:
; NAME/KEY: allele
; LOCATION: 143839
; OTHER INFORMATION: polymorphic base A or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 146668
; OTHER INFORMATION: polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 147281
; OTHER INFORMATION: polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 147505
; OTHER INFORMATION: polymorphic base G or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 148183
; OTHER INFORMATION: deletion T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 148372
; OTHER INFORMATION: polymorphic base A or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 149012
; OTHER INFORMATION: polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 149113
; OTHER INFORMATION: polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 151637
; OTHER INFORMATION: polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 151748
; OTHER INFORMATION: deletion G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 151769
; OTHER INFORMATION: polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 151847
; OTHER INFORMATION: polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 152691
; OTHER INFORMATION: polymorphic base A or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 152766
; OTHER INFORMATION: polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 153046
; OTHER INFORMATION: polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 153123
; OTHER INFORMATION: polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 153925
; OTHER INFORMATION: polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 153977
; OTHER INFORMATION: polymorphic base G or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 154502

; OTHER INFORMATION: polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 154677
; OTHER INFORMATION: polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 154879
; OTHER INFORMATION: polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 154918
; OTHER INFORMATION: polymorphic base G or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 155802
; OTHER INFORMATION: polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 156448
; OTHER INFORMATION: polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 157238
; OTHER INFORMATION: polymorphic base A or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 157897
; OTHER INFORMATION: polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 158172
; OTHER INFORMATION: polymorphic base A or G
; FEATURE:

Query Match 4.5%; Score 18; DB 5; Length 319608;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 134 GAGAAGAAGGAAAGGAGA 151
||||||||||||||||||
Db 75504 GAGAAGAAGGAAAGGAGA 75487

RESULT 3
US-09-539-333D-1/c
; Sequence 1, Application US/09539333D
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Bihain, Bernard
; APPLICANT: Essioux, Laurent
; TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
; FILE REFERENCE: GENSET.047AUS
; CURRENT APPLICATION NUMBER: US/09/539,333D
; CURRENT FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US 60/126,903
; PRIOR FILING DATE: 1999-03-30
; PRIOR APPLICATION NUMBER: US 60/131,971
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/132,065
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/143,928
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: US 60/145,915
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: US 60/146,453
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/146,452
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/162,288
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: US 09/416,384
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: Patent.pm
; SEQ ID NO 1
; LENGTH: 319608
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 31..1107
; OTHER INFORMATION: 5'regulatory region g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1108..1289
; OTHER INFORMATION: exon A g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 14877..14920
; OTHER INFORMATION: exon B g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 18778..18862
; OTHER INFORMATION: exon Bbis g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 25593..25740
; OTHER INFORMATION: exon C g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 29388..29502
; OTHER INFORMATION: exon D g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 29967..30282
; OTHER INFORMATION: exon E g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 64666..64812
; OTHER INFORMATION: exon F g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 65505..65853
; OTHER INFORMATION: exon G g35018 gene
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 65854..67854
; OTHER INFORMATION: 3'regulatory region g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 94124..94964
; OTHER INFORMATION: exon g35017
; FEATURE:
; NAME/KEY: exon
; LOCATION: 201188..201234
; OTHER INFORMATION: exon S g35030 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 214676..214793
; OTHER INFORMATION: exon T g35030 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 215702..215746
; OTHER INFORMATION: exon U g35030 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 216836..216915
; OTHER INFORMATION: exon V g35030 gene
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 213818..215818
; OTHER INFORMATION: 3'regulatory region g34872 gene

; FEATURE:
; NAME/KEY: exon
; LOCATION: 215819..215941
; OTHER INFORMATION: exon R complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 215819..215975
; OTHER INFORMATION: exon Rbis complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 216661..216952
; OTHER INFORMATION: exon Qbis complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 216661..217061
; OTHER INFORMATION: exon Q complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 217027..217061
; OTHER INFORMATION: exon Q1 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 229647..229742
; OTHER INFORMATION: exon X complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 230408..230721
; OTHER INFORMATION: exon P complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 231272..231412
; OTHER INFORMATION: exon Obis complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 231787..231880
; OTHER INFORMATION: exon O2 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 231870..231879
; OTHER INFORMATION: exon O1 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 234174..234321
; OTHER INFORMATION: exon O complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 237406..237428
; OTHER INFORMATION: exon Nbis complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 239719..239807
; OTHER INFORMATION: exon N2 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 239719..239853
; OTHER INFORMATION: exon N complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 240528..240569
; OTHER INFORMATION: exon M1117 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 240528..240596
; OTHER INFORMATION: exon M1090 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 240528..240617
; OTHER INFORMATION: exon M1069 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 240528..240644
; OTHER INFORMATION: exon MS2 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 240528..240824
; OTHER INFORMATION: exon M862 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 240528..240994
; OTHER INFORMATION: exon M692 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 240528..241685
; OTHER INFORMATION: exon M1 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 240800..240993
; OTHER INFORMATION: exon MS1 complement g34872 gene
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 241686..243685
; OTHER INFORMATION: 5'regulatory region g34872 gene
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 290652..292652
; OTHER INFORMATION: 3'regulatory region g34665 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 292653..292841
; OTHER INFORMATION: exon B complement g34665 gene

Query Match 4.5%; Score 18; DB 5; Length 319608;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 134 GAGAAGAAGGAAAGGAGA 151
||||||||||||||||||
Db 75504 GAGAAGAAGGAAAGGAGA 75487

RESULT 4
US-09-534-850-21604
; Sequence 21604, Application US/09534850
; GENERAL INFORMATION:
; APPLICANT: Seilhamer, Jeffrey J.
; APPLICANT: Delegeane, Angelo M.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Mullahy, Sara J.
; APPLICANT: Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING OR REGULATING METABOLIC MOLECULES
; FILE REFERENCE: PD-1020 CIP
; CURRENT APPLICATION NUMBER: US/09/534,850
; CURRENT FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: 07/916,491
; PRIOR FILING DATE: 1992-07-17
; PRIOR APPLICATION NUMBER: 07/977,780
; PRIOR FILING DATE: 1992-11-19
; PRIOR APPLICATION NUMBER: 08/100,523
; PRIOR FILING DATE: 1993-08-03
; PRIOR APPLICATION NUMBER: 09/008,119
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/196,364
; PRIOR FILING DATE: 1994-02-14
; PRIOR APPLICATION NUMBER: 08/282,991
; PRIOR FILING DATE: 1994-07-28
; PRIOR APPLICATION NUMBER: 08/438,571
; PRIOR FILING DATE: 1995-05-10
; PRIOR APPLICATION NUMBER: 08/179,873
; PRIOR FILING DATE: 1994-01-11
; PRIOR APPLICATION NUMBER: 08/504,732
; PRIOR FILING DATE: 1995-07-20
; PRIOR APPLICATION NUMBER: 08/137,951
; PRIOR FILING DATE: 1993-10-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 26730

```
;  SOFTWARE: PERL Program
; SEQ ID NO 21604
;   LENGTH: 236
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: hu00952490
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 210, 228
;   OTHER INFORMATION: a, t, c, g, or other
US-09-534-850-21604

  Query Match             4.2%;  Score 17;  DB 5;  Length 236;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    138 AGAAGGAAAGGAGATGT 154
          |||||||||||||||||
Db    125 AGAAGGAAAGGAGATGT 141


RESULT 5
US-09-534-850-20438
; Sequence 20438, Application US/09534850
; GENERAL INFORMATION:
;  APPLICANT: Seilhamer, Jeffrey J.
;  APPLICANT:  Delegeane, Angelo M.
;  APPLICANT:  Stuart, Susan G.
;  APPLICANT:  Stuve, Laura L.
;  APPLICANT:  Mullahy, Sara J.
;  APPLICANT:  Naughton, Rebecca E.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING OR REGULATING METABOLIC MOLECULES
;  FILE REFERENCE: PD-1020 CIP
;  CURRENT APPLICATION NUMBER: US/09/534,850
;  CURRENT FILING DATE:  2000-03-24
;  PRIOR APPLICATION NUMBER: 07/916,491
;  PRIOR FILING DATE: 1992-07-17
;  PRIOR APPLICATION NUMBER: 07/977,780
;  PRIOR FILING DATE: 1992-11-19
;  PRIOR APPLICATION NUMBER: 08/100,523
;  PRIOR FILING DATE: 1993-08-03
;  PRIOR APPLICATION NUMBER: 09/008,119
;  PRIOR FILING DATE: 1998-01-16
;  PRIOR APPLICATION NUMBER: 08/196,364
;  PRIOR FILING DATE: 1994-02-14
;  PRIOR APPLICATION NUMBER: 08/282,991
;  PRIOR FILING DATE: 1994-07-28
;  PRIOR APPLICATION NUMBER: 08/438,571
;  PRIOR FILING DATE: 1995-05-10
;  PRIOR APPLICATION NUMBER: 08/179,873
;  PRIOR FILING DATE: 1994-01-11
;  PRIOR APPLICATION NUMBER: 08/504,732
;  PRIOR FILING DATE: 1995-07-20
;  PRIOR APPLICATION NUMBER: 08/137,951
;  PRIOR FILING DATE: 1993-10-14
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 26730
;  SOFTWARE: PERL Program
; SEQ ID NO 20438
;   LENGTH: 238
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: hu00483915
US-09-534-850-20438

  Query Match             4.2%;  Score 17;  DB 5;  Length 238;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    138 AGAAGGAAAGGAGATGT 154
          |||||||||||||||||
Db     89 AGAAGGAAAGGAGATGT 105


RESULT 6
US-09-513-999C-20190/c
; Sequence 20190, Application US/09513999C
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Duclert, A.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
;  FILE REFERENCE: 59.US2.REG
;  CURRENT APPLICATION NUMBER: US/09/513,999C
;  CURRENT FILING DATE:  2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/122,487
;  PRIOR FILING DATE: 1999-02-26
;  NUMBER OF SEQ ID NOS: 36681
;  SOFTWARE: Patent.pm
; SEQ ID NO 20190
;   LENGTH: 254
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 133
;   OTHER INFORMATION: s=g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 136
;   OTHER INFORMATION: y=c or t
US-09-513-999C-20190

  Query Match             4.2%;  Score 17;  DB 5;  Length 254;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    135 AGAAGAAGGAAAGGAGA 151
          |||||||||||||||||
Db    182 AGAAGAAGGAAAGGAGA 166


RESULT 7
US-09-513-999C-14719/c
; Sequence 14719, Application US/09513999C
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Duclert, A.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
;  FILE REFERENCE: 59.US2.REG
;  CURRENT APPLICATION NUMBER: US/09/513,999C
;  CURRENT FILING DATE:  2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/122,487
;  PRIOR FILING DATE: 1999-02-26
;  NUMBER OF SEQ ID NOS: 36681
;  SOFTWARE: Patent.pm
; SEQ ID NO 14719
;   LENGTH: 261
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 254
;   OTHER INFORMATION: r=a or g
US-09-513-999C-14719

  Query Match             4.2%;  Score 17;  DB 5;  Length 261;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    135 AGAAGAAGGAAAGGAGA 151
```

```
          ||||||||||||||||
Db    171 AGAAGAAGGAAAGGAGA 155


RESULT 8
US-09-513-999C-33198
; Sequence 33198, Application US/09513999C
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Duclert, A.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
;  FILE REFERENCE: 59.US2.REG
;  CURRENT APPLICATION NUMBER: US/09/513,999C
;  CURRENT FILING DATE:  2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/122,487
;  PRIOR FILING DATE: 1999-02-26
;  NUMBER OF SEQ ID NOS: 36681
;  SOFTWARE: Patent.pm
; SEQ ID NO 33198
;   LENGTH: 324
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-513-999C-33198

  Query Match             4.2%;  Score 17;  DB 5;  Length 324;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     10 GCTCACCAGCTGGCCCA 26
          ||||||||||||||||
Db    121 GCTCACCAGCTGGCCCA 137


RESULT 9
US-10-092-411A-66
; Sequence 66, Application US/10092411A
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 032796-101
;  CURRENT APPLICATION NUMBER: US/10/092,411A
;  CURRENT FILING DATE:  2002-03-07
;  PRIOR APPLICATION NUMBER: US 09/134,001
;  PRIOR FILING DATE: 1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5676
; SEQ ID NO 66
;   LENGTH: 489
;   TYPE: DNA
;   ORGANISM: Staphylococcus epidermidis
US-10-092-411A-66

  Query Match             4.2%;  Score 17;  DB 6;  Length 489;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    223 CTTTGACTTCAGGTGGA 239
          ||||||||||||||||
Db    332 CTTTGACTTCAGGTGGA 348


RESULT 10
US-10-092-411A-1790
; Sequence 1790, Application US/10092411A
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 032796-101
;  CURRENT APPLICATION NUMBER: US/10/092,411A
;  CURRENT FILING DATE:  2002-03-07
;  PRIOR APPLICATION NUMBER: US 09/134,001
;  PRIOR FILING DATE: 1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5676
; SEQ ID NO 1790
;   LENGTH: 597
;   TYPE: DNA
;   ORGANISM: Staphylococcus epidermidis
US-10-092-411A-1790

  Query Match             4.2%;  Score 17;  DB 6;  Length 597;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    250 AGGCAATATTTAAATCT 266
          |||||||||||||||||
Db    503 AGGCAATATTTAAATCT 519


RESULT 11
US-09-724-676-25002/c
; Sequence 25002, Application US/09724676
; GENERAL INFORMATION:
;  APPLICANT: Compugen LTD
;  TITLE OF INVENTION: Variants of alternative splicing
;  FILE REFERENCE: 129181.4 Compugen
;  CURRENT APPLICATION NUMBER: US/09/724,676
;  CURRENT FILING DATE:  2000-11-28
;  NUMBER OF SEQ ID NOS: 97222
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 25002
;   LENGTH: 1165
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-724-676-25002

  Query Match             4.2%;  Score 17;  DB 5;  Length 1165;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    135 AGAAGAAGGAAAGGAGA 151
          ||||||||||||||||
Db    218 AGAAGAAGGAAAGGAGA 202


RESULT 12
US-09-724-676A-25002/c
; Sequence 25002, Application US/09724676A
; GENERAL INFORMATION:
;  APPLICANT: Compugen LTD
;  TITLE OF INVENTION: Variants of alternative splicing
;  FILE REFERENCE: 129181.4 Compugen
;  CURRENT APPLICATION NUMBER: US/09/724,676A
;  CURRENT FILING DATE:  2000-11-28
;  NUMBER OF SEQ ID NOS: 97222
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 25002
;   LENGTH: 1165
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-724-676A-25002

  Query Match             4.2%;  Score 17;  DB 5;  Length 1165;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    135 AGAAGAAGGAAAGGAGA 151
```

```
          ||||||||||||||||
Db    218 AGAAGAAGGAAAGGAGA 202


RESULT 13
US-09-724-676-25006/c
; Sequence 25006, Application US/09724676
; GENERAL INFORMATION:
;  APPLICANT: Compugen LTD
;  TITLE OF INVENTION: Variants of alternative splicing
;  FILE REFERENCE: 129181.4 Compugen
;  CURRENT APPLICATION NUMBER: US/09/724,676
;  CURRENT FILING DATE:  2000-11-28
;  NUMBER OF SEQ ID NOS: 97222
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 25006
;   LENGTH: 1199
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-724-676-25006

  Query Match             4.2%;  Score 17;  DB 5;  Length 1199;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    135 AGAAGAAGGAAAGGAGA 151
          ||||||||||||||||
Db    252 AGAAGAAGGAAAGGAGA 236


RESULT 14
US-09-724-676A-25006/c
; Sequence 25006, Application US/09724676A
; GENERAL INFORMATION:
;  APPLICANT: Compugen LTD
;  TITLE OF INVENTION: Variants of alternative splicing
;  FILE REFERENCE: 129181.4 Compugen
;  CURRENT APPLICATION NUMBER: US/09/724,676A
;  CURRENT FILING DATE:  2000-11-28
;  NUMBER OF SEQ ID NOS: 97222
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 25006
;   LENGTH: 1199
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-724-676A-25006

  Query Match             4.2%;  Score 17;  DB 5;  Length 1199;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    135 AGAAGAAGGAAAGGAGA 151
          ||||||||||||||||
Db    252 AGAAGAAGGAAAGGAGA 236


RESULT 15
US-09-724-676-25004/c
; Sequence 25004, Application US/09724676
; GENERAL INFORMATION:
;  APPLICANT: Compugen LTD
;  TITLE OF INVENTION: Variants of alternative splicing
;  FILE REFERENCE: 129181.4 Compugen
;  CURRENT APPLICATION NUMBER: US/09/724,676
;  CURRENT FILING DATE:  2000-11-28
;  NUMBER OF SEQ ID NOS: 97222
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 25004
;   LENGTH: 1203
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-724-676-25004

  Query Match             4.2%;  Score 17;  DB 5;  Length 1203;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    135 AGAAGAAGGAAAGGAGA 151
          ||||||||||||||||
Db    218 AGAAGAAGGAAAGGAGA 202


Search completed: January  2, 2003, 17:55:58
Job time : 426 secs
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: January 2, 2003, 15:04:27 ; Search time 2202 Seconds
(without alignments)
2956.669 Million cell updates/sec

Title: US-09-850-147-1
Perfect score: 402
Sequence: 1 ccgggccgggctcaccagct.....ggcataacttcctgtgttcc 402

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	29	7.2	352	12	BG354266	BG354266 947033G03
c	2	29	7.2	365	13	BI388689	BI388689 949044B01
	3	29	7.2	400	13	BI396083	BI396083 949044B01
c	4	29	7.2	479	10	AW091358	AW091358 614094A10
	5	29	7.2	487	10	BE598350	BE598350 PI1_81_E1
	6	29	7.2	493	12	BG558648	BG558648 RHIZ2_58_
	7	29	7.2	535	10	AW679158	AW679158 WS1_22_A0
c	8	29	7.2	537	12	BG842699	BG842699 MEST39-A0
	9	29	7.2	568	10	AW565695	AW565695 LG1_348_A
c	10	29	7.2	570	9	AI978092	AI978092 614022D04
c	11	29	7.2	590	10	AW288508	AW288508 618068C02
	12	29	7.2	592	10	AW036911	AW036911 614022D04
	13	29	7.2	606	14	BQ656076	BQ656076 pRC0407 S
	14	29	7.2	612	10	AW745589	AW745589 WS1_35_A0
c	15	29	7.2	613	9	AI691590	AI691590 606022D02
	16	29	7.2	614	10	BE597103	BE597103 PI1_70_G0
	17	29	7.2	681	13	BG946826	BG946826 IP1_1_B10
c	18	29	7.2	1349	11	AY109457	AY109457 Zea mays
c	19	22	5.5	356	9	AI619388	AI619388 486093F06
	20	22	5.5	574	10	BE020893	BE020893 sm53f03.y
c	21	20	5.0	284	10	BB007187	BB007187 BB007187
	22	20	5.0	300	10	BE401569	BE401569 CNW03PL46
	23	20	5.0	409	13	BM448247	BM448247 DSA023D02
	24	20	5.0	502	17	AQ821035	AQ821035 HS_4728_B
	25	20	5.0	696	17	AQ781302	AQ781302 HS_3104_B
	26	20	5.0	867	17	CNS055CG	AL321865 Tetraodon
c	27	20	5.0	928	14	BQ685675	BQ685675 AGENCOURT
c	28	19	4.7	108	9	AI891239	AI891239 614020F10
c	29	19	4.7	315	17	TA122H01P	AL462669 T. brucei
	30	19	4.7	329	10	AW946083	AW946083 QV0-ET000
	31	19	4.7	338	9	AA734352	AA734352 vv24f11.r
c	32	19	4.7	343	17	BH849501	BH849501 SALK_0697
c	33	19	4.7	347	12	BF358635	BF358635 QV0-ET000
c	34	19	4.7	349	10	AW946080	AW946080 QV0-ET000
c	35	19	4.7	354	9	AV031208	AV031208 AV031208
	36	19	4.7	357	10	AW755733	AW755733 sl08d10.y
c	37	19	4.7	358	12	BF358608	BF358608 QV0-ET000
c	38	19	4.7	363	12	BF358617	BF358617 QV0-ET000
c	39	19	4.7	366	12	BF358620	BF358620 QV0-ET000
	40	19	4.7	372	17	BH104961	BH104961 RPCI-24-2
	41	19	4.7	378	17	AZ638587	AZ638587 1M0498L05
c	42	19	4.7	413	10	AW924010	AW924010 WS1_32_C0
	43	19	4.7	414	14	BQ454260	BQ454260 sao78d08.
c	44	19	4.7	416	17	AQ280819	AQ280819 CITBI-E1-
	45	19	4.7	423	10	AW704751	AW704751 sk55a07.y

ALIGNMENTS

RESULT 1
BG354266/c
LOCUS BG354266 352 bp mRNA linear EST 05-MAR-2001
DEFINITION 947033G03.x1 947 - 2 week shoot from Barkan lab Zea mays cDNA, mRNA sequence.
ACCESSION BG354266
VERSION BG354266.1 GI:13198328
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 352)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 947033 row: G column: 03.
FEATURES Location/Qualifiers
source 1. .352
/organism="Zea mays"

/cultivar="B73"
/db_xref="taxon:4577"
/clone_lib="947 - 2 week shoot from Barkan lab"
/tissue_type="leaf and stem, including leaf base"
/dev_stage="2 week old seedling (3 leaves)"
/lab_host="XL1-Blue"
/note="Organ: shoot; Vector: Lambda ZAP (pBlueScript SK-); Site_1: EcoRI; Site_2: XhoI; Directionally cloned using Stratagene's UniZap XR cDNA cloning kit with the 5' end at the EcoRI site. The library represents 8 x 10e5 independent recombinant phage. The plants were greenhouse grown."
BASE COUNT 106 a 75 c 65 g 106 t
ORIGIN

Query Match 7.2%; Score 29; DB 12; Length 352;
Best Local Similarity 100.0%; Pred. No. 6.7e-05;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 112 AGCTTCTATATTAAGCTTCGATGAGAAGA 140
||||||||||||||||||||||||||||
Db 266 AGCTTCTATATTAAGCTTCGATGAGAAGA 238

RESULT 2
BI388689/c
LOCUS BI388689 365 bp mRNA linear EST 02-AUG-2001
DEFINITION 949044B01.x1 949 - Juvenile leaf and shoot cDNA from Steve Moose Zea mays cDNA, mRNA sequence.
ACCESSION BI388689
VERSION BI388689.1 GI:15079375
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 365)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 949044 row: B column: 01.
FEATURES Location/Qualifiers
source 1. .365
/organism="Zea mays"
/cultivar="W64A"
/db_xref="taxon:4577"
/clone_lib="949 - Juvenile leaf and shoot cDNA from Steve Moose"
/tissue_type="immature leaf primordium and vegetative meristem"
/dev_stage="4 stages from 3-13 days after imbibing"
/lab_host="E. coli XLOLR"
/note="Organ: juvenile vegetative shoots; Vector: pAD-GAL4-2.1; Site_1: EcoRI; Site_2: XhoI; Equal amounts of total RNA by weight from 4 tissue sources (see below) were pooled, polyA+ RNA isolated, and cDNA synthesized for EcoRI (5') and XhoI (3') directional cloning into lambda Hybrizap vector from Stratagene. Tissue Sources: 1. Whole shoots 3 days after sowing/imbibing in wet soil. 2. Basal 1.5 cm shoots 6 days after sowing - includes yellow portions of developing leaves 1-5, primordia from 6-8, and the vegetative apex. 3. Non-green portions of developing leaves 4-5 and the vegetative apex, including adult leaf primordia, 9 days after sowing. 4. Partially expanded and greening leaves 4-5 at 13 days after sowing."
BASE COUNT 108 a 81 c 70 g 106 t
ORIGIN

Query Match 7.2%; Score 29; DB 13; Length 365;
Best Local Similarity 100.0%; Pred. No. 6.7e-05;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 112 AGCTTCTATATTAAGCTTCGATGAGAAGA 140
||||||||||||||||||||||||||||
Db 264 AGCTTCTATATTAAGCTTCGATGAGAAGA 236

RESULT 3
BI396083
LOCUS BI396083 400 bp mRNA linear EST 08-AUG-2001
DEFINITION 949044B01.y1 949 - Juvenile leaf and shoot cDNA from Steve Moose Zea mays cDNA, mRNA sequence.
ACCESSION BI396083
VERSION BI396083.1 GI:15100292
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 400)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 949044 row: B column: 01.
FEATURES Location/Qualifiers
source 1. .400
/organism="Zea mays"
/cultivar="W64A"
/db_xref="taxon:4577"
/clone_lib="949 - Juvenile leaf and shoot cDNA from Steve Moose"
/tissue_type="immature leaf primordium and vegetative meristem"
/dev_stage="4 stages from 3-13 days after imbibing"
/lab_host="E. coli XLOLR"
/note="Organ: juvenile vegetative shoots; Vector: pAD-GAL4-2.1; Site_1: EcoRI; Site_2: XhoI; Equal amounts of total RNA by weight from 4 tissue sources (see below) were pooled, polyA+ RNA isolated, and cDNA synthesized for EcoRI (5') and XhoI (3') directional cloning into lambda Hybrizap vector from Stratagene. Tissue Sources: 1. Whole shoots 3 days after sowing/imbibing in wet soil. 2. Basal 1.5 cm shoots 6 days after sowing - includes yellow portions of developing leaves 1-5, primordia from 6-8, and the vegetative apex. 3. Non-green portions of developing leaves 4-5 and the vegetative apex, including adult leaf primordia, 9 days after sowing. 4. Partially expanded and greening leaves 4-5 at 13 days after sowing."
BASE COUNT 122 a 77 c 81 g 120 t
ORIGIN

Query Match 7.2%; Score 29; DB 13; Length 400;
Best Local Similarity 100.0%; Pred. No. 6.8e-05;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 112 AGCTTCTATATTAAGCTTCGATGAGAAGA 140
||||||||||||||||||||||||||||
Db 55 AGCTTCTATATTAAGCTTCGATGAGAAGA 83

```
RESULT 4
AW091358/c
LOCUS       AW091358                479 bp    mRNA    linear   EST 18-OCT-1999
DEFINITION  614094A10.x1 614 - root cDNA library from Walbot Lab Zea mays cDNA,
            mRNA sequence.
ACCESSION   AW091358
VERSION     AW091358.1  GI:6056953
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 479)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 614094  row: A  column: 10.
FEATURES             Location/Qualifiers
     source          1. .479
                     /organism="Zea mays"
                     /cultivar="W23"
                     /db_xref="taxon:4577"
                     /clone_lib="614 - root cDNA library from Walbot Lab"
                     /tissue_type="root"
                     /dev_stage="3-4 days old"
                     /lab_host="XLOLR"
                     /note="Organ: root; Vector: pBlueScriptII SK+; Site_1:
                     EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot
                     Lab (LM)"
BASE COUNT      142 a    111 c     96 g    130 t
ORIGIN

  Query Match             7.2%;  Score 29;  DB 10;  Length 479;
  Best Local Similarity   100.0%;  Pred. No. 7e-05;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     112 AGCTTCTATATTAAGCTTCGATGAGAAGA 140
           ||||||||||||||||||||||||||||||
Db     299 AGCTTCTATATTAAGCTTCGATGAGAAGA 271


RESULT 5
BE598350
LOCUS       BE598350                487 bp    mRNA    linear   EST 18-AUG-2000
DEFINITION  PI1_81_E10.g1_A002 Pathogen induced 1 (PI1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   BE598350
VERSION     BE598350.1  GI:9853423
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 487)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: pathogen-induced plants
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: PolyTMix
            High quality sequence start: 10
            High quality sequence stop: 477
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .487
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Pathogen induced 1 (PI1)"
                     /note="Organ: Anthracnose-infected leaves from
                     two-week-old sorghum plants 48 hr after inoculation;
                     Vector: pBluescript II from Lambda Zap II; Site_1: XhoI;
                     Site_2: EcoRI; Two-week-old sorghum plants (BTX 623
                     cultivar) were infected with pathogen (isolate FRM42I of
                     Colletotrichum graminicola, which is a sorghum isolate).
                     RNA was prepared from infected leaves harvested from 45
                     seedlings  48 hours after inoculation.  Note: young
                     seedlings (2 weeks old) exhibit juvenile resistant
                     reaction, which is an incompatible interaction. As they
                     grow older (4 weeks or older), plants resume susceptiblity
                     to anthracnose disease. The library was made from poly-A
                     RNA in the cloning vector lambda ZAP II. Clones to be
                     sequenced were prepared by mass excision. WARNING: While
                     most or all ESTs are expected to derive from the host
                     plant, no effort was made to eliminate ESTs deriving from
                     the pathogen."
BASE COUNT      132 a     94 c    125 g    136 t
ORIGIN

  Query Match             7.2%;  Score 29;  DB 10;  Length 487;
  Best Local Similarity   100.0%;  Pred. No. 7.1e-05;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     112 AGCTTCTATATTAAGCTTCGATGAGAAGA 140
           ||||||||||||||||||||||||||||||
Db     312 AGCTTCTATATTAAGCTTCGATGAGAAGA 340


RESULT 6
BG558648
LOCUS       BG558648                493 bp    mRNA    linear   EST 10-APR-2001
DEFINITION  RHIZ2_58_E10.g1_A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA
            sequence.
ACCESSION   BG558648
VERSION     BG558648.1  GI:13587646
KEYWORDS    EST.
SOURCE      Sorghum propinquum.
  ORGANISM  Sorghum propinquum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 493)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: Sorghum propinquum rhizomes
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Seq primer: T7
            High quality sequence start: 70
            High quality sequence stop: 419
```

```
          POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .493
                     /organism="Sorghum propinquum"
                     /db_xref="taxon:132711"
                     /clone_lib="Rhizome2 (RHIZ2)"
                     /note="Organ: Rhizomes; Vector: pBluescript II from Lambda
                     Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                     from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      131 a     90 c    115 g    157 t
ORIGIN

  Query Match             7.2%;  Score 29;  DB 12;  Length 493;
  Best Local Similarity   100.0%;  Pred. No. 7.1e-05;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     112 AGCTTCTATATTAAGCTTCGATGAGAAGA 140
           ||||||||||||||||||||||||||||||
Db     156 AGCTTCTATATTAAGCTTCGATGAGAAGA 184


RESULT 7
AW679158
LOCUS       AW679158                535 bp    mRNA    linear   EST 19-JUL-2000
DEFINITION  WS1_22_A09.g1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   AW679158
VERSION     AW679158.1  GI:7552912
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 535)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: water-stressed plants
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: T7
            High quality sequence start: 48
            High quality sequence stop: 523
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .535
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Water-stressed 1 (WS1)"
                     /note="Organ: Mix of 5-week old plants on days 7 & 8 after
                     water was withheld; Vector: Lambda Zap; Site_1: XhoI;
                     Site_2: EcoRI; The library was made from poly-A RNA in the
                     cloning vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT      137 a     91 c    131 g    176 t
ORIGIN

  Query Match             7.2%;  Score 29;  DB 10;  Length 535;
  Best Local Similarity   100.0%;  Pred. No. 7.2e-05;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     112 AGCTTCTATATTAAGCTTCGATGAGAAGA 140
           ||||||||||||||||||||||||||||||
Db     164 AGCTTCTATATTAAGCTTCGATGAGAAGA 192


RESULT 8
BG842699/c
LOCUS       BG842699                537 bp    mRNA    linear   EST 29-MAY-2001
DEFINITION  MEST39-A09.T3 ISUM4-TN Zea mays cDNA clone MEST39-A09 3', mRNA
            sequence.
ACCESSION   BG842699
VERSION     BG842699.2  GI:14244761
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 537)
  AUTHORS   Qiu,F., Cui,F., Guo,L., Ashlock,D.A, Wen,T.J. and Schnable,P.S.
  TITLE     Expressed Sequence Tags from B73 Maize Seedlings and Silks
  JOURNAL   Unpublished (2001)
COMMENT     On May 25, 2001 this sequence version replaced gi:14209021.
            Contact: Patrick S. Schnable
            Schnable Laboratory
            Iowa State University
            G405 Agronomy, Iowa State University, Ames, IA 50011-1010, USA
            Tel: 515-294-0975
            Fax: 515-294-2299
            Email: schnable@iastate.edu
            PCR PRimers
            FORWARD: T7-1 (AA TAC GAC TCA CTA TAG)
            BACKWARD: T3 (ATT AAC CCT CAC TAA AG)
            Seq primer: primer T3 (ATT AAC CCT CAC TAA AG).
FEATURES             Location/Qualifiers
     source          1. .537
                     /organism="Zea mays"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="MEST39-A09"
                     /clone_lib="ISUM4-TN"
                     /tissue_type="Seedling and silk"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3PAC; Site_1: EcoRI; Site_2: NotI;
                     ds-cDNA molecules were generated as follows. First-strand
                     cDNA was prepared from oligo-dT selected mRNA by priming
                     with a NotI  oligo-dT primer (5'
                     AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTT).  The
                     resulting DNA:RNA hybrid was treated with RNase H and used
                     as a  template for DNA PolI-catalyzed second strand
                     synthesis. After the  addition of EcoRI adaptors, the
                     ds-cDNAs were digested with NotI and  size-selected. The
                     resulting molecules were directionally cloned  into the
                     EcoRI and NotI sites of the pT7T3PAC vector. The library
                     then went through one round of normalization to CoT value
                     of 5 based  on the methods of Marcelo Bento Soares (Genome
                     Research 6: 791-806,  1996)."
BASE COUNT      156 a    115 c     97 g    169 t
ORIGIN

  Query Match             7.2%;  Score 29;  DB 12;  Length 537;
  Best Local Similarity   100.0%;  Pred. No. 7.2e-05;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     112 AGCTTCTATATTAAGCTTCGATGAGAAGA 140
           ||||||||||||||||||||||||||||||
Db     353 AGCTTCTATATTAAGCTTCGATGAGAAGA 325


RESULT 9
AW565695
LOCUS       AW565695                568 bp    mRNA    linear   EST 19-JUL-2000
DEFINITION  LG1_348_A05.g1_A002 Light Grown 1 (LG1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   AW565695
```

```
VERSION     AW565695.1  GI:7219573
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 568)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C. and Pratt,L.H.
  TITLE     An EST database from Sorghum: light-grown seedlings
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: T7
            High quality sequence start: 29
            High quality sequence stop: 568
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .568
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Light Grown 1 (LG1)"
                     /note="Organ: 10- to 14-day-old light-grown (greenhouse)
                     seedlings; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI
                     ; The library was made from poly-A RNA in the cloning
                     vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT      147 a     99 c    142 g    180 t
ORIGIN

  Query Match             7.2%;  Score 29;  DB 10;  Length 568;
  Best Local Similarity   100.0%;  Pred. No. 7.3e-05;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     112 AGCTTCTATATTAAGCTTCGATGAGAAGA 140
           ||||||||||||||||||||||||||||||
Db     225 AGCTTCTATATTAAGCTTCGATGAGAAGA 253


RESULT 10
AI978092/c
LOCUS       AI978092                570 bp    mRNA    linear   EST 27-AUG-1999
DEFINITION  614022D04.x2 614 - root cDNA library from Walbot Lab Zea mays cDNA,
            mRNA sequence.
ACCESSION   AI978092
VERSION     AI978092.1  GI:5791300
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 614022  row: D  column: 04.
FEATURES             Location/Qualifiers
     source          1. .570
                     /organism="Zea mays"
                     /cultivar="W23"
                     /db_xref="taxon:4577"
                     /clone_lib="614 - root cDNA library from Walbot Lab"
                     /tissue_type="root"
                     /dev_stage="3-4 days old"
                     /lab_host="XLOLR"
                     /note="Organ: root; Vector: pBlueScriptII SK+; Site_1:
                     EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot
                     Lab (LM)"
BASE COUNT      177 a    133 c    112 g    148 t
ORIGIN

  Query Match             7.2%;  Score 29;  DB 9;  Length 570;
  Best Local Similarity   100.0%;  Pred. No. 7.3e-05;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     112 AGCTTCTATATTAAGCTTCGATGAGAAGA 140
           ||||||||||||||||||||||||||||||
Db     391 AGCTTCTATATTAAGCTTCGATGAGAAGA 363


RESULT 11
AW288508/c
LOCUS       AW288508                590 bp    mRNA    linear   EST 16-JAN-2000
DEFINITION  618068C02.x2 618 - Inbred Tassel cDNA Library Zea mays cDNA, mRNA
            sequence.
ACCESSION   AW288508
VERSION     AW288508.1  GI:6695400
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 590)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 618068  row: C  column: 02.
FEATURES             Location/Qualifiers
     source          1. .590
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="618 - Inbred Tassel cDNA Library"
                     /tissue_type="tassel"
                     /dev_stage="tassel length from 0.1 to 2.5 cm"
                     /lab_host="XLOLR"
                     /note="Organ: tassel; Vector: pAD-GAL4-2.1 (Hybrizap);
                     Inbred tassel library from Schmidt lab"
BASE COUNT      182 a    138 c    112 g    157 t      1 others
ORIGIN

  Query Match             7.2%;  Score 29;  DB 10;  Length 590;
  Best Local Similarity   100.0%;  Pred. No. 7.3e-05;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     112 AGCTTCTATATTAAGCTTCGATGAGAAGA 140
           ||||||||||||||||||||||||||||||
Db     330 AGCTTCTATATTAAGCTTCGATGAGAAGA 302
```

```
RESULT 12
AW036911
LOCUS       AW036911                592 bp    mRNA    linear   EST 15-SEP-1999
DEFINITION  614022D04.y1 614 - root cDNA library from Walbot Lab Zea mays cDNA,
            mRNA sequence.
ACCESSION   AW036911
VERSION     AW036911.1  GI:5895665
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 592)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 614022  row: D  column: 04.
FEATURES             Location/Qualifiers
     source          1. .592
                     /organism="Zea mays"
                     /cultivar="W23"
                     /db_xref="taxon:4577"
                     /clone_lib="614 - root cDNA library from Walbot Lab"
                     /tissue_type="root"
                     /dev_stage="3-4 days old"
                     /lab_host="XLOLR"
                     /note="Organ: root; Vector: pBlueScriptII SK+; Site_1:
                     EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot
                     Lab (LM)"
BASE COUNT      159 a    120 c    142 g    171 t
ORIGIN

  Query Match             7.2%;  Score 29;  DB 10;  Length 592;
  Best Local Similarity   100.0%;  Pred. No. 7.3e-05;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     112 AGCTTCTATATTAAGCTTCGATGAGAAGA 140
           ||||||||||||||||||||||||||||||
Db     305 AGCTTCTATATTAAGCTTCGATGAGAAGA 333


RESULT 13
BQ656076
LOCUS       BQ656076                606 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  pRC0407 Sorghum halepense rhizome cDNA library Sorghum halepense
            cDNA clone pRC0407, mRNA sequence.
ACCESSION   BQ656076
VERSION     BQ656076.1  GI:21788402
KEYWORDS    EST.
SOURCE      Sorghum halepense.
  ORGANISM  Sorghum halepense
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 606)
  AUTHORS   Paterson,A.H.
  TITLE     Unpublished, Paterson,A.H
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Paterson AH
            Center for Applied Genetic Technologies
            University of Georgia
            Riverbend Research Laboratory, Room 162, 110 Riverbend Road, Athens
            , GA 30602, USA
            Tel: 7065830162
            Fax: 7065830160
            Email: paterson@uga.edu
            contig
            Seq primer: M13F/M13R.
FEATURES             Location/Qualifiers
     source          1. .606
                     /organism="Sorghum halepense"
                     /db_xref="taxon:4560"
                     /clone="pRC0407"
                     /clone_lib="Sorghum halepense rhizome cDNA library"
                     /tissue_type="rhizome"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                     XhoI; cDNA library was prepared from mRNA isolated from
                     the apical 2-3 cm of actively growing rhizomes of Johnson
                     grass (Sorghum halepense) and cloned into the EcoRI/XhoI
                     sites of lambda ZAP II (Stratagene)."
BASE COUNT      152 a    110 c    147 g    197 t
ORIGIN

  Query Match             7.2%;  Score 29;  DB 14;  Length 606;
  Best Local Similarity   100.0%;  Pred. No. 7.3e-05;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     112 AGCTTCTATATTAAGCTTCGATGAGAAGA 140
           ||||||||||||||||||||||||||||||
Db     196 AGCTTCTATATTAAGCTTCGATGAGAAGA 224


RESULT 14
AW745589
LOCUS       AW745589                612 bp    mRNA    linear   EST 19-JUL-2000
DEFINITION  WS1_35_A09.g1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   AW745589
VERSION     AW745589.1  GI:7659327
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 612)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: water-stressed plants
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: T7
            High quality sequence start: 7
            High quality sequence stop: 612
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .612
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Water-stressed 1 (WS1)"
                     /note="Organ: Mix of 5-week old plants on days 7 & 8 after
                     water was withheld; Vector: Lambda Zap; Site_1: XhoI;
                     Site_2: EcoRI; The library was made from poly-A RNA in the
                     cloning vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT      156 a    110 c    148 g    198 t
ORIGIN
```

```
  Query Match             7.2%;  Score 29;  DB 10;  Length 612;
  Best Local Similarity   100.0%;  Pred. No. 7.4e-05;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     112 AGCTTCTATATTAAGCTTCGATGAGAAGA 140
           ||||||||||||||||||||||||||||||
Db     203 AGCTTCTATATTAAGCTTCGATGAGAAGA 231


RESULT 15
AI691590/c
LOCUS       AI691590                613 bp    mRNA    linear   EST 02-FEB-2000
DEFINITION  606022D02.x1 606 - Ear tissue cDNA library from Schmidt lab Zea
            mays cDNA, mRNA sequence.
ACCESSION   AI691590
VERSION     AI691590.1  GI:4966734
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 613)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 606022  row: D  column: 02.
FEATURES             Location/Qualifiers
     source          1. .613
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="606 - Ear tissue cDNA library from Schmidt
                     lab"
                     /tissue_type="mixed"
                     /dev_stage="ear length from 0.5 cm - 2.0 cm"
                     /lab_host="XLOLR (Stratagene)"
                     /note="Organ: immature ear; Vector: pBK-CMV; Site_1: EcoRI
                     ; Site_2: XhoI; Mixed ear tissue cDNA library from Schmidt
                     lab"
BASE COUNT      178 a    138 c    114 g    183 t
ORIGIN

  Query Match             7.2%;  Score 29;  DB 9;  Length 613;
  Best Local Similarity   100.0%;  Pred. No. 7.4e-05;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     112 AGCTTCTATATTAAGCTTCGATGAGAAGA 140
           ||||||||||||||||||||||||||||||
Db     348 AGCTTCTATATTAAGCTTCGATGAGAAGA 320


Search completed: January  2, 2003, 16:44:11
Job time : 2207 secs
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: January 2, 2003, 12:21:06 ; Search time 3182 Seconds
(without alignments)
3676.722 Million cell updates/sec

Title: US-09-850-147-1
Perfect score: 402
Sequence: 1 ccgggccgggctcaccagct.....ggcataacttcctgtgttcc 402

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		113.2	28.2	170226	8	AP003450	AP003450 Oryza sat
2		96.8	24.1	912	8	AY097414	AY097414 Arabidops
3		96.8	24.1	1068	8	AF275664	AF275664 Arabidops
4		96.8	24.1	1440	8	AY064136	AY064136 Arabidops
5		91.4	22.7	1312	8	AY087422	AY087422 Arabidops
6		91.4	22.7	1366	8	AF428374	AF428374 Arabidops
7		86.8	21.6	1213	8	AF305635	AF305635 Pisum sat
8		82.4	20.5	1324	8	MDPPMD1	Z47076 M.domestica
9		43.6	10.8	31522	8	AP002050	AP002050 Arabidops
10	c	42.8	10.6	100975	8	AC007980	AC007980 Arabidops
11	c	42.4	10.5	108992	8	AP003740	AP003740 Oryza sat
12	c	42.4	10.5	142081	8	AP004572	AP004572 Oryza sat
13	c	40.8	10.1	158294	8	AC091724	AC091724 Oryza sat
14		40.4	10.0	936	6	E22116	E22116 Yeast capab
15		40.4	10.0	1978	8	YSCSIT4C	M24395 Saccharomyc
16		40.4	10.0	3536	8	SCYDL047W	Z74095 S.cerevisia
17		40.4	10.0	36687	8	SCCIVL37K	Z71781 S.cerevisia
18		40	10.0	165577	2	AC095138	AC095138 Rattus no
19		39.8	9.9	196679	2	AC130153	AC130153 Rattus no
20	c	39.2	9.8	149131	2	AC119135	AC119135 Rattus no
21	c	38.2	9.5	129135	9	AL136162	AL136162 Human DNA
22	c	38.2	9.5	194716	2	AL136227	AL136227 Homo sapi
23	c	38	9.5	39520	8	SPBC26H8	AL031743 S.pombe c
24	c	37.8	9.4	177939	9	AC087616	AC087616 Homo sapi
25	c	37.8	9.4	183279	2	AC079014	AC079014 Homo sapi
26		37.8	9.4	211464	9	AC023950	AC023950 Homo sapi
27		37.4	9.3	39355	2	AC130364	AC130364 Homo sapi
28		37	9.2	625	4	AF273675	AF273675 Nyctalus
29	c	37	9.2	1037	11	G67201	G67201 Xq4125 KWOK
30	c	37	9.2	100625	9	HS1189K21	AL030997 Human DNA
31		36.6	9.1	544	6	AX186776	AX186776 Sequence
32	c	36.6	9.1	173285	2	AC118963	AC118963 Rattus no
33	c	36.6	9.1	224297	2	AC121768	AC121768 Mus muscu
34	c	36.4	9.1	1971	6	AX299314	AX299314 Sequence
35	c	36.4	9.1	2529	6	AX375232	AX375232 Sequence
36	c	36.4	9.1	2553	6	AX283620	AX283620 Sequence
37	c	36.4	9.1	58198	9	AC073524	AC073524 Homo sapi
38		36.4	9.1	111103	9	AC007129	AC007129 Homo sapi
39		36.4	9.1	113991	9	AC096742	AC096742 Homo sapi
40		36.4	9.1	154733	2	AC127462	AC127462 Danio rer
41	c	36.4	9.1	164232	2	AC067792	AC067792 Homo sapi
42	c	36.4	9.1	174830	9	AL805977	AL805977 Human DNA
43		36.4	9.1	188209	2	AC099649	AC099649 Homo sapi
44	c	36.2	9.0	123116	2	AL360234	AL360234 Homo sapi
45	c	36.2	9.0	167611	9	AL583859	AL583859 Human DNA

ALIGNMENTS

RESULT 1
AP003450
LOCUS AP003450 170226 bp DNA linear PLN 21-MAR-2002
DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, PAC clone:P0034C09.
ACCESSION AP003450
VERSION AP003450.3 GI:16904687
KEYWORDS .
SOURCE Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA, clone:P0034C09.
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1

```
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
            clone:P0034C09
  JOURNAL   Published Only in Database (2001)
REFERENCE   2  (bases 1 to 170226)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-2001) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     On Nov 12, 2001 this sequence version replaced gi:15887058.
            Genes were predicted from the integrated results of the following:
            GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor
            (October 1998 version). The genomic sequence was searched against
            NCBI NonRedundant Protein database, nr
            (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
            RGP. Protein homologies of the coding regions were searched against
            NCBI NonRedundant Protein database with BLASTP2.0. ESTs represent
            the identified cDNA sequences using BLASTN 2.0 with the
            corresponding DDBJ accession no. and RGP clone ID.
            A gene with identity or significant homology to a protein is
            classified based on the protein name to indicate the homology level
            such as same name, 'putative-' and '-like protein'.  A gene without
            significant homology to any protein but with EST homology (covering
            almost the entire length of partial sequence) is classified as an
            'unknown' protein. A gene predicted with a gene prediction program
            is classified as a 'hypothetical' protein.
            The orientation of the sequence is from SP6 to T7 of the PAC clone.
            This sequence of P0034C09 clone has an overlap with P0519D04 (DDBJ:
            AP003455) at 5' end and with P0431H09(DDBJ: AP003248) at 3' end.
            Detailed information on overlap and assembly quality together with
            annotation of this entry is available at
            http://rgp.dna.affrc.go.jp/GenomeSeq.html.
FEATURES             Location/Qualifiers
     source          1. .170226
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="1"
                     /clone="P0034C09"
     gene            complement(join(2898. .3354,3514. .3964,4062. .4343,
                     6999. .7872))
                     /gene="P0034C09.1"
     CDS             complement(join(2898. .3354,3514. .3964,4062. .4343,
                     6999. .7872))
                     /gene="P0034C09.1"
                     /note="contains ESTs
                     AU032946(S00005),C22642(S15001),D48660(S15001)
                     unknown protein"
                     /codon_start=1
                     /protein_id="BAB84588.1"
                     /db_xref="GI:18565401"
                     /translation="MSRLCHHGGRVLLPLLLVAAACLGDPTGDTYDTAMCGAQASITC
                     GGVIVRYPFYLSNATRALPKYANSSTFCGYPGLEIICDGGGGGKAVMMLGNDSYTVSR
                     IDYASLTVSLADADVANGTCPVVSHNVTIPPAPSSLHLADTVGMLIFFFRCAFGPAAN
                     APPKPPSIHPLTCGENSEDAPTQSFLLPASPLPPGDLWHRGCSAVYGVPVLGGSLPSD
                     ANDPAWRKDGYIASLRKGFQMSWDRSDRCSRCELTSGKCGYNQNGKFLGCLCANGLVD
                     SDACSKISDSTLRLAGVVGGGLSAVFALGLIATVFFVRKRKHKKVNSSSKLLKYSGSG
                     GTPRSMGGDMESGSVKDLQTHLFSYEELEEATDSFNENRELGDGGFGTVYKGILRDGR
                     VVAVKRLYNNSYRRVEQFVNEAAILSRLRHPNLVMFYGCTSSQSRELLLVYEFVANGT
                     VADHLHGHRAQERALSWPLRLNIAVESAAALTYLHAIEPPIVHRDVKTTNILLDADFH
                     VKVADFGLSRLFPLDVTHVSTAPQGTPGYVDPEYHQCYQLTDKSDVYSFGVVLVELIS
                     SKPAVDITRQRNEINLAGMAINRIQKSQLEELVDLELGYESDPATKKMMTMVAELAFR
                     CLQQNGEMRPPIKEVLEGLKGVQDLCVMEKDGGKDKKGPDPPLSPDTVHAQWDSRQTT
                     PNTSQ"
     gene            complement(11635. .12537)
                     /gene="P0034C09.2"
     CDS             complement(11635. .12537)
                     /gene="P0034C09.2"
                     /note="contains EST C26543(C12558)
                     unknown protein"
                     /codon_start=1
                     /protein_id="BAB84589.1"
                     /db_xref="GI:18565402"
                     /translation="MPLLLCRRRRRLLPLLLVLVAASHGASNDTYDTSMCLQETTTCG
                     DVSIRYPFYFSDKTGYINGSSNSYCGYPGLAIDCDDGKPILQLNGAEKYKVNYINVGS
                     ITNVSLVDKEVVDDSSGCPRVDHNVTFAQGSWLFFPAGMSLDYLVFFLGCSFPNLFLP
                     PENIDPITCSFIGLIGPSYVLPKDQVPPGNWSQFCKTFEVPVVKYQQMDPKGDAWRKG
                     GYGQVLRQGFLLSVNDSRRPPNCTQCEESKGRCGFSQDGEFIGCLCLNGRVRSLRCGS
                     SDLTTGKPFVSRFALYLQRPKGLI"
     gene            complement(14885. .15808)
                     /gene="P0034C09.3"
     CDS             complement(14885. .15808)
                     /gene="P0034C09.3"
                     /note="contains ESTs AU100785(C50325),C26880(C50325)
                     unknown protein"
                     /codon_start=1
                     /protein_id="BAB84590.1"
                     /db_xref="GI:18565403"
                     /translation="MAHLPLLLLSFLLLIGVHASVSDGSPLPHTYNTSICSKSYKCGG
                     VNISYPFYLSNATGETYDYTQFSCGYTDLNITCSWDGSKQTPFIQLNGDNYTILEIIY
                     DSRTIVLVDTDALRGGSCPRVRHNITFSQADEWLQYTGPRDNLTFFFGCNLVPLPPMD
                     PGLTRFVDKNQINCKDFSNWPDGGDSFVFTSVELEAPMEYELASRCRQVIVVPVNGSI
                     LNSSSQSALPSGEYGQVLKKGFELAWKSRKDKQCNLCEQSRGQCAYSQNRTFLGCLCA
                     DGKVSSTDCTSSKSFDADPSFFSPPFCKANN"
     gene            complement(21989. .22951)
                     /gene="P0034C09.4"
     CDS             complement(21989. .22951)
                     /gene="P0034C09.4"
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="BAB84591.1"
                     /db_xref="GI:18565404"
                     /translation="MPPPLLLLPVVLALFAAADASPSACRNATCGDVPVAYPFWLIDN
                     SSALGCGYPGLGLRCVDNTTLILPFRTHQYRVLDIEYGKHTISLTDADMQHSDTNNSC
                     PHLHANLTIDDNSWLQLASSDSNITFLYNCKSNSSLLSSTSVVKLTGCGAGPEHIGSS
                     YVFLDGWITGEAYDYECETVVVAPVIDAHKKAMVDAHGGLPPADGSFGDVLSAGFELT
                     YSAHSDKCGKCERTKGWCGYRHNETSTTMDFTCFCDEGPTKSHCGTHASSPLTMSLLL
                     FGVLLASDFRIGAVDIEFELLVERLKMINKPIRIINFSKQLLLH"
     gene            join(23815. .23822,24395. .24764)
                     /gene="P0034C09.5"
     CDS             join(23815. .23822,24395. .24764)
                     /gene="P0034C09.5"
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="BAB84592.1"
                     /db_xref="GI:18565405"
                     /translation="MRSGGRRDDGHAQMRAAGGARRQSSGGRARQRSGDGRQAERTAS
                     EEAGTMTSGVTEMRATMASASSSAGAAFPRQTGHEVWDESHMLMHSWWNTCPQAEICL
                     TFSPSSTARRQTTHLNDAAVSEI"
     gene            complement(join(25181. .25819,25829. .26579,26587. .26790,
                     26825. .26964))
                     /gene="P0034C09.6"
     CDS             complement(join(25181. .25819,25829. .26579,26587. .26790,
                     26825. .26964))
                     /gene="P0034C09.6"
                     /codon_start=1
                     /product="similar to Oryza sativa chromosome 1,
                     P0031E09.21"
                     /protein_id="BAB84593.1"
                     /db_xref="GI:18565406"
                     /translation="MDEGCILDFFGRQGRGPRIEGGSAEPFGPARARERRPIRARGRG
                     RERWGPLVGDTAHGGMGAREVRRKGFGCAWLRSGAPGGRMAAGPRAAARLVVDRAHGR
                     EDGGRGRLTGLARRRGANVAATRAGGRRKKGRPRWTADGGRVHRSKRNGFGLRGAHRS
                     DEGVEKERANGSDSQEGVRRWWNSTAAAGGERKGKSPRGDEDSIKGDWSISGVQGTHF
                     RGRMGLSGAVEVRRREAASGGGRQWRWGHAGWRRSLGRRLRAFGKATELPRKLEGTRV
                     SREGEQEGGTTEPGRRGDEQPTARKRALRRRAGEGEEDGRPESRVTRPRTHWLPRCSA
                     MDGDGASRAEATLARWRRERKRERLGDGAARRWGMSVLGGRKPTSWAVKEGRSRRTWR
                     QGVVTGGRGGMVPVGEKAKGEGGKGALPMPFREKGGRREDSASGSWTHAAWNGRARVA
                     GGCGAERRRRGRAAWAAAWRRLPGLTRRRPGSWPRVGERARDAGGGGVLPDFERRRRE
                     IGKAAVHSAARGERERKQRESEERERKKRGDASLSLGVCSARAGRRRSGREERMGRAR
                     EKRPTAKRGAI"
     gene            complement(28952. .29269)
                     /gene="P0034C09.7"
```

```
     CDS             complement(28952. .29269)
                     /gene="P0034C09.7"
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="BAB84594.1"
                     /db_xref="GI:18565407"
                     /translation="MMLFGALAASFALSFFLITIFICPRALRIACRRRDRPLVMEQEQ
                     QQPTLPRFGLDTATITRLPSFIYVTPQSGTAVQRHLTGACRGKNGANRFLRNRDLSTS
                     PRT"
     gene            join(32474. .32478,33007. .33659,33836. .33861)
                     /gene="P0034C09.8"
     CDS             join(32474. .32478,33007. .33659,33836. .33861)
                     /gene="P0034C09.8"
                     /note="contains ESTs C26880(C50325),AU100785(C50325)
                     unknown protein"
                     /codon_start=1
                     /protein_id="BAB84595.1"
                     /db_xref="GI:18565408"
                     /translation="MDLDGSKQTPFIQLNGENYTILEIIYDSRTIVLADTDALRGSCP
                     RVRHNVTFGQAYPWLQYTGSRDNLTFFFGCKLNLPPPIDPGLVSLADKHQINCKDFSN
                     WPDSGDSFVFTSGELEAPVESELARCCSQVIVVPVNGSVLNSSNQSALPSGGYGQVLK
                     MGFDLAWNSSKDEQCYKCEQSKGHCSYSQNRAFLGCLCSDGKVSTKDCRNNGASNSSA
                     LNNIFAVQM"
     gene            join(34381. .35201,36446. .37107,37314. .37375)
                     /gene="P0034C09.9"
     CDS             join(34381. .35201,36446. .37107,37314. .37375)
                     /gene="P0034C09.9"
                     /note="contains ESTs C26494(C12436),AU057321(S21346)"
                     /codon_start=1
                     /product="receptor serine/threonine kinase like protein"
                     /protein_id="BAB84596.1"
                     /db_xref="GI:18565409"
                     /translation="MLLSLCLRLLPLLLVLVAASHGDASGDTYDTSMCLQKPTTCGNV
                     SISYPFYFATKTKDINGSSNSYCGYPGLAIDCDDGKPILQLNGTEKYKVNYINYGSIT
                     NVSLADLEVVDDSSGCPRVDHNVTIPQISWLFFSGISVDYLVFFLRCSFTTFAPKPAN
                     FNPIACGSFINLTRPSFVFPDELVPPGNWSQLCEETFEVPVLKYQLMEMDSNGNAWNN
                     SGYAQVLRQGFQLSVNESRRPPNCSQCEESQGRCGYSQAGEFIGCLCPNGRVRSLRGN
                     LSDGRQVAVKMLKDSKGDGEEFINEVASISRTSHVNVVTLLGFCLHGSKRVLIYEYMP
                     NGSLERYAFRNNSEGEHSLTWEKLFDVVVGIARGLEYLHRGCNTRIVHFDIKPHNILL
                     DQEFCPKISDFGMAKLCSNKESIISIAGARGTIGYIAPEVYSKQFGAISSKSDVYSYG
                     MMILEMVGARERNIDANSESSSHYFPQWIYEHLDEYCINSSEIDVLYNCKYYLAVTKC

  Query Match             28.2%;  Score 113.2;  DB 8;  Length 170226;
  Best Local Similarity   73.2%;  Pred. No. 4.1e-22;
  Matches  186;  Conservative    0;  Mismatches   63;  Indels    5;  Gaps    3;

Qy    143 GAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAACCAGATTCCAGGGCCAAG 202
          | || |||| |||||  |||||||||||||| ||||| || |||||||| | ||| || ||
Db  79830 GGAAAGAGAGGTCAAATTCTTCACAGAGACAGAGGAGAATAACCAGATGCGAGGACCGAG 79889

Qy    203 GGCTTGGAATCCCATATTTCCTTTGACT---TCAGGTGGACAATTTTAGAAGGCAATATT 259
          | | ||   |||||||||||||| |||||   |||  || |||||||||||||| | || ||
Db  79890 GAC-TGCTGTCCCATATTTCCTCTGACTTCATCAATTGTACAATTTTAGAAGTC-ATGTT 79947

Qy    260 TAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGGAGGTTCAGATCATAA 319
          |  ||||| ||||| ||||||  || ||||     ||| |  || |||| ||| |||||
Db  79948 TTTATCTGTAAGGGTTAAAGATGAAGCAGACCTATTTATTGTTGTAGGTCCAGGTCATAG 80007

Qy    320 TGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACAAGTACAAAAAACCAACGGG 379
          ||| |||| |||| |   |||| |   |||||||| | |||| |||||| | ||||   ||
Db  80008 TGCAATCTAATGCGTATGATTTTTCTGTTCTTTTAACCACATGTACAATAGACCATTTGG 80067

Qy    380 AGGGGCATAACTTC 393
          || | ||| ||| |
Db  80068 AGTGTCATCACTCC 80081


RESULT 2
AY097414
LOCUS       AY097414                912 bp    mRNA    linear   PLN 05-MAY-2002
DEFINITION  Arabidopsis thaliana AT3g19980/MZE19_3 mRNA, complete cds.
ACCESSION   AY097414
VERSION     AY097414.1  GI:20453382
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 912)
  AUTHORS   Cheuk,R., Chen,H., Kim,C.J., Meyers,M.C., Shinn,P., Banh,J.,
            Bowser,L., Carninci,P., Chang,E., Dale,J.M., Goldsmith,A.D.,
            Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G.,
            Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M.,
            Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T.,
            Satou,M., Seki,M., Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C.,
            Yamada,K., Yamamura,Y., Yu,G., Yu,S., Shinozaki,K., Davis,R.W.,
            Theologis,A. and Ecker,J.R.
  TITLE     Arabidopsis ORF clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 912)
  AUTHORS   Cheuk,R., Chen,H., Kim,C.J., Meyers,M.C., Shinn,P., Banh,J.,
            Bowser,L., Carninci,P., Chang,E., Dale,J.M., Goldsmith,A.D.,
            Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G.,
            Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M.,
            Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T.,
            Satou,M., Seki,M., Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C.,
            Yamada,K., Yamamura,Y., Yu,G., Yu,S., Shinozaki,K., Davis,R.W.,
            Theologis,A. and Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-APR-2002) Salk Institute Genomic Analysis Laboratory
            (SIGnAL), Plant Biology Laboratory, The Salk Institute for
            Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
            USA
COMMENT     RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-Length cDNA') :  Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P.,  Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members constructed and
            sequenced the pUNI (ORF) clones using the RAFL cDNAs: Cheuk,R.,
            Chen,H., Kim,C.J., Meyers,M.C., Shinn,P., Banh,J. Bowser,L.,
            Chan,M.M., Chang,E., Dale,J.M., Deng, J.M., Goldsmith,A.D.,
            Jones,T., Karlin-Neumann,G., Lam,B., Lee,J.M., Lin,J., Miranda,M.,
            Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Southwick,A.,
            Tang,C.C., Toriumi,M., Wu,H.C., Yamada,K., Yamamura, Y., Yu,G.,
            Davis,R.W.,  Theologis,A., and Ecker,J.R.

            Cheuk,R. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to
            this work.  Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk)
            contributed equally to this work as PIs.
FEATURES             Location/Qualifiers
     source          1. .912
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="3"
                     /clone="U21104"
                     /note="This clone is in pUNI 51
                     ecotype: Columbia"
     CDS             1. .912
                     /note="serine/threonine protein phosphatase, putative"
                     /codon_start=1
                     /product="AT3g19980/MZE19_3"
                     /protein_id="AAM19930.1"
                     /db_xref="GI:20453383"
                     /translation="MDLDQWISKVKDGQHLSEDELQLLCEYVKEILIEESNVQPVNSP
                     VTVCGDIHGQFHDLMKLFQTGGHVPDTNYIFMGDFVDRGYNSLEVFTILLLLKARYPA
                     NITLLRGNHESRQLTQVYGFYDECQRKYGNANAWRYCTDVFDYLTLSAIIDGTVLCVH
                     GGLSPDVRTIDQIRLIERNCEIPHEGPFCDLMWSDPEDIETWAVSPRGAGWLFGSRVT
                     TEFNHINKLDLVCRAHQLVQEGLKYMFQDKGLVTVWSAPNYCYRCGNVASILSFNDNM
                     EREVKFFTETEENNQMRGPRTGVPYFL"
BASE COUNT      250 a    181 c    221 g    260 t
ORIGIN

  Query Match             24.1%;  Score 96.8;  DB 8;  Length 912;
```

```
  Best Local Similarity   73.2%;  Pred. No. 1.9e-17;
  Matches  164;  Conservative    0;  Mismatches   57;  Indels    3;  Gaps    3;

Qy       5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
           |||| || ||||| || |  ||   | | |||||  |||||| ||| || ||  ||||||
Db     692 GCCGCGCGCACCAACTTGTACAAGAAGGTCTTAAGTACATGTTCCAAGATAA-AGGCCTT 750

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
           | |  ||| ||||||||||  | |  |    | | ||||| ||||  ||||||||||
Db     751 GTAACTGTATGGTCTGCACCTAATTACTGTTACCGCTGTGGGAATGTCGCTTCTATATTG 810

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
           || ||| |||| || | ||||||| || || ||| ||||||||||||| || || |||||
Db     811 AGTTTCAATGACAACATGGAAAGG-GAAGTGAAGTTCTTCACAGAGACAGAAGAGAACAA 869

Qy     185 CCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGA 228
           || ||   |||||||||| | |||| | || |||||||| |||
Db     870 TCAAATGAGAGGGCCAAGGAC-TGGAGTTCCGTATTTCCTATGA 912


RESULT 3
AF275664
LOCUS       AF275664               1068 bp    mRNA    linear   PLN 02-JUL-2001
DEFINITION  Arabidopsis thaliana serine/threonine protein phosphatase (STPP)
            mRNA, complete cds.
ACCESSION   AF275664
VERSION     AF275664.1  GI:14582205
KEYWORDS    .
SOURCE      Arabidopsis thaliana.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1068)
  AUTHORS   Kang,J.G., Kim,D.-H. and Park,C.-M.
  TITLE     Isolation of cDNA clone encoding a putative serine/threonine
            protein phosphatase
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1068)
  AUTHORS   Kang,J.G., Kim,D.-H. and Park,C.-M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUN-2000) Kumho Life & Environmental Science
            Laboratory, 1 Oryong-Dong, Puk-Gu, Kwangju 500-712, Korea
FEATURES             Location/Qualifiers
     source          1. .1068
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
     gene            1. .1068
                     /gene="STPP"
     CDS             98. .1009
                     /gene="STPP"
                     /codon_start=1
                     /product="serine/threonine protein phosphatase"
                     /protein_id="AAK69404.1"
                     /db_xref="GI:14582206"
                     /translation="MDLDQWISKVKDGQHLSEDELQLLCEYVKEILIEESNVQPVNSP
                     VTVCGDIHGQFHDLMKLFQTGGHVPDTNYIFMGDFVDRGYNSLEVFTILLLLKARYPA
                     NITLLRGNHESRQLTQVYGFYDECQRKYGNANAWRYCTDVFDYLTLSAIIDGTVLCVH
                     GGLSPDVRTIDQIRLIERNCEIPHEGPFCDLMWSDPEDIETWAVSPRGAGWLFGSRVT
                     TEFNHINKLDLVCRAHQLVQEGLKYMFQDKGLVTVWSAPNYCYRCGNVASILSFNDNM
                     EREVKFFTETEENNQMRGPRTGVPYFL"
BASE COUNT      299 a    205 c    253 g    311 t
ORIGIN

  Query Match             24.1%;  Score 96.8;  DB 8;  Length 1068;
  Best Local Similarity   73.2%;  Pred. No. 1.9e-17;
  Matches  164;  Conservative    0;  Mismatches   57;  Indels    3;  Gaps    3;

Qy       5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
           |||| || ||||| || |  ||   | | |||||  |||||| ||| || ||  ||||||
Db     789 GCCGCGCGCACCAACTTGTACAAGAAGGTCTTAAGTACATGTTCCAAGATAA-AGGCCTT 847

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
           | |  ||| ||||||||||  | |  |    | | ||||| ||||  ||||||||||
Db     848 GTAACTGTATGGTCTGCACCTAATTACTGTTACCGCTGTGGGAATGTCGCTTCTATATTG 907

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
           || ||| |||| || | ||||||| || || ||| ||||||||||||| || || |||||
Db     908 AGTTTCAATGACAACATGGAAAGG-GAAGTGAAGTTCTTCACAGAGACAGAAGAGAACAA 966

Qy     185 CCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGA 228
           || ||   |||||||||| | |||| | || |||||||| |||
Db     967 TCAAATGAGAGGGCCAAGGAC-TGGAGTTCCGTATTTCCTATGA 1009


RESULT 4
AY064136
LOCUS       AY064136               1440 bp    mRNA    linear   PLN 05-DEC-2001
DEFINITION  Arabidopsis thaliana AT3g19980/MZE19_3 mRNA, complete cds.
ACCESSION   AY064136
VERSION     AY064136.1  GI:17381248
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1440)
  AUTHORS   Shinn,P., Chen,H., Cheuk,R., Kim,C.J., Koesema,E., Meyers,M.C.,
            Banh,J., Bowser,L., Carninci,P., Chang,E., Dale,J.M.,
            Goldsmith,A.D., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A.,
            Karlin-Neumann,G., Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M.,
            Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L.,
            Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C., Toriumi,M.,
            Wu,H.C., Yamada,K., Yamamura,Y., Yu,G., Yu,S., Shinozaki,K.,
            Davis,R.W., Theologis,A. and Ecker,J.R.
  TITLE     Arabidopsis cDNA clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1440)
  AUTHORS   Shinn,P., Chen,H., Cheuk,R., Kim,C.J., Koesema,E., Meyers,M.C.,
            Banh,J., Bowser,L., Carninci,P., Chang,E., Dale,J.M.,
            Goldsmith,A.D., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A.,
            Karlin-Neumann,G., Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M.,
            Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L.,
            Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C., Toriumi,M.,
            Wu,H.C., Yamada,K., Yamamura,Y., Yu,G., Yu,S., Shinozaki,K.,
            Davis,R.W., Theologis,A. and Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-NOV-2001) Salk Institute Genomic Analysis Laboratory
            (SIGnAL), Plant Biology Laboratory, The Salk Institute for
            Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
            USA
COMMENT     RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-Length cDNA') :  Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P.,  Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members carried out the
            sequencing and annotation of the RAFL cDNAs: Shinn,P., Chen,H.,
            Cheuk,R., Kim,C.J., Koesema,E., Meyers,M.C., Banh,J. Bowser,L.,
            Chang,E., Dale,J.M., Goldsmith,A.D., Jones,T., Karlin-Neumann,G.,
            Lam,B., Lee,J.M., Lin,J., Miranda,M., Nguyen,M., Onodera,C.S.,
            Palm,C.J., Quach,H.L., Southwick,A., Tang,C.C., Toriumi,M.,
            Wu,H.C., Yamada,K., Yamamura, Y., Yu,G., Yu,S., Davis,R.W.,
            Theologis,A., and Ecker,J.R.

            Shinn,P. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to
            this work.  Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk)
            contributed equally to this work as PIs.
FEATURES             Location/Qualifiers
     source          1. .1440
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="3"
                     /clone="RAFL09-21-A14(R21104)"
```

```
                     /note="ecotype: Columbia"
     CDS             184. .1095
                     /note="serine/threonine protein phosphatase, putative"
                     /codon_start=1
                     /product="AT3g19980/MZE19_3"
                     /protein_id="AAL36043.1"
                     /db_xref="GI:17381249"
                     /translation="MDLDQWISKVKDGQHLSEDELQLLCEYVKEILIEESNVQPVNSP
                     VTVCGDIHGQFHDLMKLFQTGGHVPDTNYIFMGDFVDRGYNSLEVFTILLLLKARYPA
                     NITLLRGNHESRQLTQVYGFYDECQRKYGNANAWRYCTDVFDYLTLSAIIDGTVLCVH
                     GGLSPDVRTIDQIRLIERNCEIPHEGPFCDLMWSDPEDIETWAVSPRGAGWLFGSRVT
                     TEFNHINKLDLVCRAHQLVQEGLKYMFQDKGLVTVWSAPNYCYRCGNVASILSFNDNM
                     EREVKFFTETEENNQMRGPRTGVPYFL"
BASE COUNT      385 a    277 c    304 g    474 t
ORIGIN

  Query Match             24.1%;  Score 96.8;  DB 8;  Length 1440;
  Best Local Similarity   73.2%;  Pred. No. 2e-17;
  Matches  164;  Conservative    0;  Mismatches   57;  Indels    3;  Gaps    3;

Qy       5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
             |||| || ||||| || |  ||   | | |||||  |||||| ||| || ||  ||||||
Db     875 GCCGCGCGCACCAACTTGTACAAGAAGGTCTTAAGTACATGTTCCAAGATAA-AGGCCTT 933

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
           | |  ||| ||||||||||  | |  |    | | ||||| ||||  ||||||||||
Db     934 GTAACTGTATGGTCTGCACCTAATTACTGTTACCGCTGTGGGAATGTCGCTTCTATATTG 993

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
           || ||| |||| || | |||||||  || || ||| ||||||||||||| || || |||||
Db     994 AGTTTCAATGACAACATGGAAAGG-GAAGTGAAGTTCTTCACAGAGACAGAAGAGAACAA 1052

Qy     185 CCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGA 228
            || ||   |||||||||| | |||| | || ||||||||  |||
Db    1053 TCAAATGAGAGGGCCAAGGAC-TGGAGTTCCGTATTTCCTATGA 1095


RESULT 5
AY087422
LOCUS       AY087422               1312 bp    mRNA    linear   PLN 26-JUN-2002
DEFINITION  Arabidopsis thaliana clone 35230 mRNA, complete sequence.
ACCESSION   AY087422
VERSION     AY087422.1  GI:21406146
KEYWORDS    FLI_CDNA.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1312)
  AUTHORS   Haas,B.J., Volfovsky,N., Town,C.D., Troukhan,M., Alexandrov,N.,
            Feldmann,K.A., Flavell,R.B., White,O. and Salzberg,S.L.
  TITLE     Full-length messenger RNA sequences greatly improve genome
            annotation
  JOURNAL   Genome Biol. (2002) In press
REFERENCE   2  (bases 1 to 1312)
  AUTHORS   Brover,V., Troukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and
            Feldmann,K.
  TITLE     Full-Length cDNA from Arabidopsis thaliana
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1312)
  AUTHORS   Brover,V., Troukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and
            Feldmann,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road,
            Malibu, CA 90265, USA
COMMENT     This clone sequence is one of 5,000 Ceres full-length cDNAs made
            available to TIGR and Genbank. The following quality assessment of
            this set was done by comparison with known proteins: two percent of
            the clones are estimated to be 5'-truncated; less than one percent
            are 3'-truncated; approximately two percent represent alternative
            splice variants, including unspliced introns and spliced exons; one
            percent may contain premature stop codons; five percent may have
            frame shifts in a coding region. A sequence is considered to be
            5'-truncated if it lacks the translation initiation start (ATG). A
            sequence is considered to be 3'-truncated if it lacks the
            C-terminal end of the encoded protein. Please note that these cDNA
            sequences are derived from the Ws or LAer ecotypes and therefore
            may contain polymorphisms when compared to sequences from Col-0.
            Genset carried out the library production and sequencing of the
            full-length clones. Ceres, Inc. carried out the clustering of the
            5' sequences, selection of clones, and sequence assembly.
FEATURES              Location/Qualifiers
     source           1. .1312
                      /organism="Arabidopsis thaliana"
                      /db_xref="taxon:3702"
                      /clone="35230"
     CDS              179. .1090
                      /codon_start=1
                      /product="phosphoprotein phosphatase"
                      /protein_id="AAM64970.1"
                      /db_xref="GI:21593021"
                      /translation="MDLDQWISKVKDGQHLSEDELQLLCEYVKEILIEESNVQPVNSP
                      VTVCGDIHGQFHDLMKLFQTGGHVPETNYIFMGDFVDRGYNSLEVFTILLLLKARHPA
                      NITLLRGNHESRQLTQVYGFYDECQRKYGNANAWRYCTDVFDYLTLSAIIDGTVLCVH
                      GGLSPDVRTIDQIRLIERNCEIPHEGPFCDLMWSDPEDIETWAVSPRGAGWLFGSRVT
                      TEFNHINNLDLVCRAHQLVQEGLKYMFQDKGLVTVWSAPNYCYRCGNVASILSFNDNM
                      EREVKFFTETEENNQMRGPRTGVPYFL"
BASE COUNT      353 a    264 c    278 g    417 t
ORIGIN

  Query Match             22.7%;  Score 91.4;  DB 8;  Length 1312;
  Best Local Similarity   70.4%;  Pred. No. 7.5e-16;
  Matches  164;  Conservative    0;  Mismatches   66;  Indels    3;  Gaps    3;

Qy       5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
           | || |||||||||||| |  ||   | | || ||  |||||| ||| || ||  || |||
Db     870 GTCGCGCTCACCAGCTTGTACAAGAAGGTCTCAAGTACATGTTCCAAGATAA-AGGTCTT 928

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
           | |  ||||||||||||| ||  | |  |    | | ||||| ||||  ||||||||||
Db     929 GTAACTGTGTGGTCTGCGCCTAATTACTGCTACCGTTGTGGGAATGTGGCTTCTATATTG 988

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
           || ||| |||| || | ||||||| || || ||| |||||||  ||||| || || |||||
Db     989 AGTTTCAATGACAACATGGAAAGG-GAAGTGAAGTTCTTCACGGAGACAGAAGAGAACAA 1047

Qy     185 CCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTG 237
            || ||   |||||||||| |  ||  | || |||||||| ||| |   | ||
Db    1048 TCAAATGAGAGGGCCAAGGAC-AGGTGTTCCTTATTTCCTGTGATTCATGATG 1099


RESULT 6
AF428374
LOCUS       AF428374               1366 bp    mRNA    linear   PLN 18-OCT-2001
DEFINITION  Arabidopsis thaliana At1g50370/F14I3_10 mRNA, complete cds.
ACCESSION   AF428374
VERSION     AF428374.1  GI:16226939
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1366)
  AUTHORS   Cheuk,R., Chen,H., Kim,C.J., Koesema,E., Meyers,M.C., Banh,J.,
            Bowser,L., Carninci,P., Dale,J.M., Goldsmith,A.D., Hayashizaki,Y.,
            Ishida,J., Jiang,P.X., Jones,T., Kamiya,A., Karlin-Neumann,G.,
            Kawai,J., Lam,B., Lee,J.M., Lin,J., Liu,S.X., Miranda,M.,
            Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Pham,P.K.,
            Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C.,
            Toriumi,M., Yamada,K., Yamamura,Y., Yu,G., Yu,S., Shinozaki,K.,
            Davis,R.W., Theologis,A. and Ecker,J.R.
  TITLE     Arabidopsis cDNA clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1366)
```

```
  AUTHORS   Cheuk,R., Chen,H., Kim,C.J., Koesema,E., Meyers,M.C., Banh,J.,
            Bowser,L., Carninci,P., Dale,J.M., Goldsmith,A.D., Hayashizaki,Y.,
            Ishida,J., Jiang,P.X., Jones,T., Kamiya,A., Karlin-Neumann,G.,
            Kawai,J., Lam,B., Lee,J.M., Lin,J., Liu,S.X., Miranda,M.,
            Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Pham,P.K.,
            Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C.,
            Toriumi,M., Yamada,K., Yamamura,Y., Yu,G., Yu,S., Shinozaki,K.,
            Davis,R.W., Theologis,A. and Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Salk Institute Genomic Analysis Laboratory
            (SIGnAL), Plant Biology Laboratory, The Salk Institute for
            Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
            USA
COMMENT     RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-Length cDNA') :  Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P.,  Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members carried out the
            sequencing and annotation of the RAFL cDNAs: Cheuk,R., Chen,H.,
            Kim,C.J., Koesema,E., Meyers,M.C., Shinn,P., Banh,J. Bowser,L.,
            Dale,J.M., Goldsmith,A.D., Jiang,P.X., Jones,T., Karlin-Neumann,G.,
            Lam,B., Lee,J.M., Lin,J., Liu,S.X., Miranda,M., Nguyen,M.,
            Onodera,C.S., Palm,C.J., Pham,P.K., Quach,H.L., Southwick,A.,
            Tang,C.C., Toriumi,M., Yamada,K., Yamamura, Y., Yu,G., Yu,S.,
            Davis,R.W.,  Theologis,A., and Ecker,J.R.

            Cheuk,R. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to
            this work.  Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk)
            contributed equally to this work as PIs.
FEATURES             Location/Qualifiers
     source          1. .1366
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="1"
                     /clone="RAFL09-68-L23(R14796)"
                     /note="ecotype: Columbia"
     CDS             162. .1073
                     /note="phosphoprotein phosphatase"
                     /codon_start=1
                     /product="At1g50370/F14I3_10"
                     /protéin_id="AAL16304.1"
                     /db_xref="GI:16226940"
                     /translation="MDLDQWISKVKDGQHLSEDELQLLCEYVKEILIEESNVQPVNSP
                     VTVCGDIHGQFHDLMKLFQTGGHVPETNYIFMGDFVDRGYNSLEVFTILLLLKARHPA
                     NITLLRGNHESRQLTQVYGFYDECQRKYGNANAWRYCTDVFDYLTLSAIIDGTVLCVH
                     GGLSPDVRTIDQIRLIERNCEIPHEGPFCDLMWSDPEDIETWAVSPRGAGWLFGSRVT
                     TEFNHINNLDLVCRAHQLVQEGLKYMFQDKGLVTVWSAPNYCYRCGNVASILSFNDNM
                     EREVKFFTETEENNQMRGPRTGVPYFL"
BASE COUNT      365 a    265 c    290 g    446 t
ORIGIN
```

```
  Query Match             22.7%;  Score 91.4;  DB 8;  Length 1366;
  Best Local Similarity   70.4%;  Pred. No. 7.5e-16;
  Matches  164;  Conservative    0;  Mismatches   66;  Indels    3;  Gaps    3;

Qy       5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
           | || |||||||||| |  ||   | | || ||  |||||| ||| || ||  || |||
Db     853 GTCGCGCTCACCAGCTTGTACAAGAAGGTCTCAAGTACATGTTCCAAGATAA-AGGTCTT 911

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
             | |  |||||||||||| ||  | |  |    | | ||||| |||| ||||||||||
Db     912 GTAACTGTGTGGTCTGCGCCTAATTACTGCTACCGTTGTGGGAATGTGGCTTCTATATTG 971

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
           || ||| |||| || | ||||||| || || ||| ||||||| ||||| || || |||||
Db     972 AGTTTCAATGACAACATGGAAAGG-GAAGTGAAGTTCTTCACGGAGACAGAAGAGAACAA 1030

Qy     185 CCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTG 237
             || ||   |||||||||| |  ||  | || |||||||| ||| |   | ||
Db    1031 TCAAATGAGAGGGCCAAGGAC-AGGTGTTCCTTATTTCCTGTGATTCATGATG 1082
```

```
RESULT 7
AF305635
LOCUS       AF305635               1213 bp    mRNA    linear   PLN 02-MAY-2002
DEFINITION  Pisum sativum serine/threonine protein phosphatase 2A mRNA,
            complete cds.
ACCESSION   AF305635
VERSION     AF305635.1  GI:20385062
KEYWORDS    .
SOURCE      Pisum sativum.
  ORGANISM  Pisum sativum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Vicieae;
            Pisum.
REFERENCE   1  (bases 1 to 1213)
  AUTHORS   Kang,J.-G., Lim,S.-H., Kim,D.-H. and Park,C.-M.
  TITLE     Identification of the cDNA clone encoding serine/threonine
            phosphatase 2A protein in pea
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1213)
  AUTHORS   Kang,J.-G., Lim,S.-H., Kim,D.-H. and Park,C.-M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-SEP-2000) Kumho Life & Environmental Science
            Laboratory, 1 Oryong-dong, Puk-gu, Kwangju 500-712, South Korea
FEATURES             Location/Qualifiers
     source          1. .1213
                     /organism="Pisum sativum"
                     /db_xref="taxon:3888"
     CDS             95. .1006
                     /note="PsPP2A; phytochrome A-interacting protein"
                     /codon_start=1
                     /product="serine/threonine protein phosphatase 2A"
                     /protein_id="AAM21172.1"
                     /db_xref="GI:20385063"
                     /translation="MDLDQWISKVKDGQHLLEDELQLLCEYVKEILIEESNVQPVNSP
                     VTVCGDIHGQFHDLMKLFQTGGHVPETNYIFMGDFVDRGYNSLEVFTILLLLKARYPA
                     NITLLRGNHESRQLTQVYGFYDECQRKYGNANAWRYCTDVFDYLTLSAIIDGTVLCVH
                     GGLSPDIRTIDQIRVIERNCEIPHEGPFCDLMWSDPEDIETWAVSPRGAGWLFGSRVT
                     SEFNHINNLDLVCRAHQLVQEGLKYMFQDKGLVTVWSAPNYCYRCGNVASILSFNENM
                     EREVKFFTETEENNQMRGPRTGVPYFL"
BASE COUNT      345 a    214 c    259 g    395 t
ORIGIN
```

```
  Query Match             21.6%;  Score 86.8;  DB 8;  Length 1213;
  Best Local Similarity   69.5%;  Pred. No. 1.7e-14;
  Matches  146;  Conservative    0;  Mismatches   62;  Indels    2;  Gaps    2;

Qy       5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
           | || || || || || | ||   | | |||||| |||||| ||| || ||  ||||||
Db     786 GTCGAGCACATCAACTTGTACAAGAAGGCCTTAAATACATGTTCCAAGATAA-AGGCCTT 844

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
             | |  ||| ||||||||||||  | |  |    | | ||||| |||| |||||||||  |
Db     845 GTAACTGTATGGTCTGCACCTAATTACTGTTACCGTTGTGGAAATGTAGCTTCTATTCTT 904

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
           || ||| |||| || | ||| | |||| || ||| | ||||| || || ||||| |||||
Db     905 AGTTTCAATGAAAATATGGAGA-GAGAAGTTAAGTTTTTCACTGAAACAGAGGAGAACAA 963

Qy     185 CCAGATTCCAGGGCCAAGGGCTTGGAATCC 214
             |||||   |||||||||| |  |   |||
Db     964 TCAGATGAGAGGGCCAAGGACAGGTGTTCC 993


RESULT 8
MDPPMD1
LOCUS       MDPPMD1                1324 bp    mRNA    linear   PLN 31-DEC-1995
DEFINITION  M.domestica Borkh mRNA for serine/threonine protein phosphatase
            (PPX).
ACCESSION   Z47076
VERSION     Z47076.1  GI:1143510
KEYWORDS    serine/threonine protein phosphatase.
```

```
SOURCE      Malus x domestica.
  ORGANISM  Malus x domestica
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Rosales; Rosaceae; Maloideae; Malus.
REFERENCE   1  (bases 1 to 1324)
  AUTHORS   Watillon,B., Puhringer,H., Kettmann,R., Boxus,P. and Burny,A.
  TITLE     Expression of PPX and PP2A phosphatase messengers in apple
  JOURNAL   Plant Sci. (1995) In press
REFERENCE   2  (bases 1 to 1324)
  AUTHORS   Watillon,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-DEC-1994) Watillon B., Faculte des Sciences
            Agronomiques, Avenue Marechal Juin, 13, GEMBLOUX, Belgium
FEATURES             Location/Qualifiers
     source          1. .1324
                     /organism="Malus x domestica"
                     /strain="Borkh"
                     /cultivar="McIntosh Wijcik"
                     /db_xref="taxon:3750"
     5'UTR           1. .47
     CDS             48. .959
                     /standard_name="PPX"
                     /codon_start=1
                     /product="Ser/Thr protein phosphatase homologous to PPX"
                     /protein_id="CAA87385.1"
                     /db_xref="GI:1143511"
                     /db_xref="SPTREMBL:Q42912"
                     /translation="MDLDQWISKVKEGQHLLEDELQLLCEYVKEILIEESNVQPVNSP
                     VTVCGDIHGQFHDLMKLFQTGGHVPETNYIFMGDFVDRGYNSLEVFTILLLLKARYPA
                     NITLLRGNHESRQLTQVYGFYDECQRKYGNANAWRYCTDVFDYLTLSAIIDGTVLCVH
                     GGLSPDIRTIDQIRVIERNCEIPHEGPFCDLMWSDPEDIETWAVSPRGAGWLFGSRVT
                     SEFNHINNLDLVCRAHQLVQEGLKYMFQDKGLVTVWSAPNYCYRCGNVASILSFNDNM
                     EREVKFFTETEENNQMRGPRTGVPYFL"
     3'UTR           960. .1324
BASE COUNT      366 a    234 c    313 g    411 t
ORIGIN

  Query Match           20.5%;  Score 82.4;  DB 8;  Length 1324;
  Best Local Similarity   69.2%;  Pred. No. 3.3e-13;
  Matches  155;  Conservative    0;  Mismatches   66;  Indels    3;  Gaps    3;

Qy       5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
           | || || ||||| || |  ||   | | |||||  ||||||  || || ||  ||||||
Db     739 GTCGAGCACACCAACTTGTTCAAGAAGGTCTTAAGTACATGTTTCAAGATAA-AGGCCTT 797

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
           | |  ||| ||||||||||  | |  |      | ||||| ||||  |||||||| ||
Db     798 GTAACTGTCTGGTCTGCACCAAATTACTGTTATCGCTGTGGGAATGTGGCTTCTATCTTG 857

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
           || ||  |||| || | ||| | |||| || ||| ||||||| || || ||||| |||||
Db     858 AGTTTTAATGACAATATGGAGA-GAGAGGTGAAGTTCTTCACTGAAACAGAGGAGAACAA 916

Qy     185 CCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGA 228
           ||||||   ||| || ||| |  ||| | || |||||| | |||
Db     917 CCAGATGAGAGGACCGAGGAC-AGGAGTACCTTATTTCTTATGA 959


RESULT 9
AP002050
LOCUS       AP002050               31522 bp    DNA     linear   PLN 27-DEC-2000
DEFINITION  Arabidopsis thaliana genomic DNA, chromosome 3, P1 clone: MZE19.
ACCESSION   AP002050 BA000014
VERSION     AP002050.1  GI:8051658
KEYWORDS    .
SOURCE      Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
            clone:MZE19.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,T., Katoh,T., Sato,S., Nakamura,A., Asamizu,E. and Tabata,S.
  TITLE     Structural analysis of Arabidopsis thaliana chromosome 3. II.
            Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC
            and BAC clones
  JOURNAL   DNA Res. 7 (3), 217-221 (2000)
  MEDLINE   20363099
REFERENCE   2  (bases 1 to 31522)
  AUTHORS   Kaneko,T., Kato,T., Sato,S., Nakamura,Y., Asamizu,E. and Tabata,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2000) Yasukazu Nakamura, Kazusa DNA Research
            Institute, Department of Plant Gene Research; 1532-3, Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
            Tel:81-438-52-3935, Fax:81-438-52-3934)
COMMENT     Address for correspondence: kaos@kazusa.or.jp
            For the latest information on annotation of this clone, please see
            http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MZE19
            Genes with similarity to proteins in the databases are described in
            'product' or 'note' qualifiers.  Genes that have no significant
            protein similarity are described as 'unknown protein'.
            The software programs used to predict genes include: Grail
            (Informatics Group, Oak Ridge National Laboratory,
            http://compbio.ornl.gov/Grail-1.3/),
            GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
            NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
            Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
            SplicePredictor (Volker Brendel, Stanford University,
            http://gremlin1.zool.iastate.edu/cgi-bin/sp.cgi).
            Genes encoding tRNAs are predicted by tRNAscan-SE
            (Sean Eddy, Washington University School of Medicine, St. Louis,
            http://genome.wustl.edu/eddy/tRNAscan-SE/).
            This sequence may not be the entire insert of this clone. It may be
            shorter because we remove overlaps between neighboring submissions.
            The 5' clone is MPN9 and the 3' clone is MAL21.
FEATURES             Location/Qualifiers
     source          1. .31522
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="3"
                     /clone="MZE19"
                     /clone_lib="Mitsui P1"
     CDS             join(1527. .1865,2003. .2104,2187. .2330,2455. .2605,
                     2715. .2742,2990. .3155,3270. .3333,3437. .3450,3552. .3636,
                     4013. .4116,4218. .4356,4652. .4770,5097. .5249,5447. .5536,
                     5673. .5750,5911. .6069,6192. .6398,6627. .6832,6920. .7055,
                     7133. .7262,7375. .7521,7626. .7693,7805. .8476)
                     /note="gene_id:MZE19.1
                     pir||S33812"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="myosin-like protein"
                     /protein_id="BAB03161.1"
                     /db_xref="GI:11994771"
                     /translation="MSQKVTPFMQSLKSLPADYRFDGSPVSDRLENSSGASVRLTNSN
                     VPRKGGLRNGVSRTDTAAGDSEDSPYSGHGVFVEEQSLTDDVDSGAATMPLPQSDERR
                     WSDTSAYARKKILQSWIQLPNGNWELGKILSTSGEESVISLPEGKVIKVISETLVPAN
                     PDILDGVDDLMQLSYLNEPSVLYNLNYRYNQDMIYTKAGPVLVAVNPFKEVPLYGNRY
                     IEAYRKKSNESPHVYAIADTAIREMIRDEVNQSIIISGESGAGKTETAKIAMQYLAAL
                     GGGSGIEYEILKTNPILEAFGNAKTLRNDNSSRFGKLIEIHFSESGKISGAQIQTFLL
                     EKSRVVQCAEGERSYHIFYQLCAGASPALREKLNLTSAHEYKYLGQSNCYSINGVDDA
                     ERFHTVKEALDIVHVSKEDQESVFAMLAAVLWLGNVSFTVIDNENHVEPVADESLSTV
                     AKLIGCNINELTLTLSKRNMRVRNDTIVQKLTLPQAIDARDALAKSIYSCLFDWLVEQ
                     INKSLAVGKRRTGRSISILDIYGFESFDKNSFEQFCINYANERLQQHFNRHLFKLEQE
                     EYIQDGIDWTRVDFEDNQNCLSLFEKKPLGLLSLLDEESTFPNGTDLTLANKLKQHLQ
                     SNSCFRGDKGKLFTVVHYAGEVTYETTGFLEKNRDLLHSDSIQLLSSCSCLLPQAFAS
                     SMLIQSEKPVVGPLYKAGGADSQRLSVATKFKSQLFQLMQRLGNTTPHFIRCIKPNNI
                     QSPGVYEQGLVLQQLRCCGVLEVVRISRSGFPTRMSHQKFSRRYGFLLVENIADRDPL
                     SVSVAILHQFNILPEMYQVGYTKLFFRTGQIGVLEDTRNRTLHGILRVQSSFRGYQAR
                     CLLKELKRGISILQSFVRGEKIRKEFAELRRRHKAAATIQSQVKSKIARIQYKGIADA
                     SVVIQSAIRGWLVRRCSGDIGWLKSGGAKTNELGEVLVKASVLSELQRRVLKAEAALR
                     EKEEEENDILQQRLQQYENRWSEYETKMKSMEEIWQKQMRSLQSSLSIAKKSLAVEDSA
                     RNSDASVNASDATDWDSSSNQFRSQTSNGVGSRLQPMSAGLSVIGRLAEEFEQRAQVF
                     GDDAKFLVEVKSGQVEANLDPDRELRRLKQMFETWKKDYGGRLRETKLILSKLGSEES
```

```
                SGSMEKVKRKWWGRRNSTRY"
CDS             join(11384. .12120,12209. .12314,12486. .12676,12837. .13107)
                /note="gene_id:MZE19.2
                unknown protein"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="BAB03162.1"
                /db_xref="GI:11994772"
                /translation="MGSLSGIIQRPLVAAAAVITASVSADVSERFSSLRSLVRGSESE
                QIAPSVSGLVQEERSLWVSQVSASKLADLSFVSRICVPVPNVDLLASNPSCSLATSVT
                SLSALRSVYQSAELAKASKPVAFTIGASLVVPDISYRWHLPEPSAVDMSGSSSCVSEK
                NRTVVVLLGWLGSKQKHLKKYADWYTSKGYHVITFTLPMNEIMSYQVGGKAEKNIESL
                VNHLADWLDEEQKKNLVFHTFSNTGWLTYGAILEKFQKQDSSLMGRVKGCIVDSAPVA
                AADPTVWASGFSAAFLKKSSVATKGSASSSYESNGINISQPKPAATETALLLVLEKFF
                AVILNLPKVNRRLADVLDTLSTAQPSCPQLYIYSSADRVIPVEQVESFIVEQRKAGHE
                VRACNFISSPHVDHFRSNPELYTAELNHFMNNFVLSCCHHSS"
CDS             join(13748. .13828,14115. .14261,14349. .14416,14658. .14715,
                15204. .15314,15398. .15448,15554. .15697,15788. .15883,
                15977. .16048,16418. .16501)
                /note="gb|AAD50050.1
                gene_id:MZE19.3"
                /codon_start=1
                /evidence=not_experimental
                /product="phosphoprotein phosphatase"
                /protein_id="BAB03163.1"
                /db_xref="GI:11994773"
                /translation="MDLDQWISKVKDGQHLSEDELQLLCEYVKEILIEESNVQPVNSP
                VTVCGDIHGQFHDLMKLFQTGGHVPDTNYIFMGDFVDRGYNSLEVFTILLLLKARYPA
                NITLLRGNHESRQLTQVYGFYDECQRKYGNANAWRYCTDVFDYLTLSAIIDGTVLCVH
                GGLSPDVRTIDQIRLIERNCEIPHEGPFCDLMWSDPEDIETWAVSPRGAGWLFGSRVT
                TEFNHINKLDLVCRAHQLVQEGLKYMFQDKGLVTVWSAPNYCYRCGNVASILSFNDNM
                EREVKFFTETEENNQMRGPRTGVPYFL"
CDS             join(17336. .18319,18474. .18842)
                /note="gene_id:MZE19.4
                unknown protein"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="BAB03164.1"
                /db_xref="GI:11994774"
                /translation="MLIKRFNPPCWRSDHHQPKESNRDKMGSSSSPSPNDAVLESARP
                FLRGELEKIDPKLPSLIAVLKSVGAGECWHKHGSFLDHLIDIYKILKLWKAPESVCLC
                GLFHSAYSNSYVNLAIFDPSTGRDVVRDHVGEAAESLIHLFCVVPRQTLIHDELLFKY
                SDCELVEHLDCSEVSLRNAKEKGIFDGDEEWRKKINALVPENGVVVKHIKTGEEIVVS
                RRVVGVFLLMTMADFSDQLFGFQDELFCNHDGRLEFRGNNVTALWPGNGKPGLWMNSN
                SRMGAIYSLIVREEEILMEERRRAFGSGFEVRKERDEDIDLVVPPVFSFCTKVLDAKE
                QIEAREMYWEVVSSDTSKEGYLERAEERLLGCIEKNPFVGEPHVLLSQVYLGKKRFKE
                AEREAEKGLLLLLQWGSPWDKRMSWEGWIAWVRVLLMKSQDQSWPDVSWGILNLGLVR
                "
CDS             join(19425. .19551,20049. .20167,20213. .20317,20601. .20651,
                20764. .20829,21160. .21261,21395. .21463,21562. .21645,
                21752. .21847,21934. .21987)
                /note="gene_id:MZE19.5"
                /codon_start=1
                /evidence=not_experimental
                /product="membrane import protein-like"
                /protein_id="BAB03165.1"
                /db_xref="GI:11994775"
                /translation="MADLLPPLTAAQVDAKTKVDEKVDYSNLPSPVPYEELHREALMS
                LKSDNFEGLRFDFTRALNQKFSLSHRYVACIYSYCCTISVMMGPTEVPAQSPETTIKI
                PTAHYEFGANYYDPKITNEEHMSQAMFNFDYMGSDYRAQLQLGQSALIGATYIQSVTN
                HLSLGGEIFWAGVPRKSGIGYAARYETDKMVASGQVASTGAVVMNYVQKISDKVSLAT
                DFMYNYFSRDVTASVGYDYMLRQARVRGKIDSNGVASALLEERLSMGLNFLLSAELDH
                KKKDYKFGFGLTVG"
CDS             join(23092. .23197,23303. .23979,24508. .24579,24654. .24731,
                24830. .25808,26191. .26321,26566. .26787,26911. .27474,
                27600. .27809,27976. .28080)
                /note="gene_id:MZE19.6"
                /codon_start=1
                /evidence=not_experimental
                /product="transcription factor-like protein"
                /protein_id="BAB03166.1"
                /db_xref="GI:11994776"
                /translation="MDSAIEISSGSDSDDEVPPQPVWPQTRTRMDPTWLSRRPLPTVD
                SHARAEHTNQAPPNGASSDTSRPGVSKPFTGNGNTVNSRISSGSGADYVRLSSEQALK
                RTLPPSFNSPPLPARSGTNNISNASGSRVGVDYERPLSQQALKRTLPPSFNPPPLPSR
                SGTNNIRNAGGSRFGADYSHPAVSAVGNKSTFGDHYSGAHAEIGIQRGVNGVRILPPS
                LTHGTSASVLHHAGSSDPMHRFGGGEDRNPDNDERLVYQAALQVLNQPMTESDLPPGT
                LSVPLMRHQKIALAWMFQKETSSFNCPGGILADDQGLGKTVSTIALILKQKIVSQLKS
                ESSCKQETEALVLDADDESDNAKHESGSHVKPELKVSSNSETSVLSACGNDENDSSDM
                EKAEDEEANSSTRAFQWKRPAAGTLIVCPASVVRQWARELDEKVSEESKLSVLVYHGS
                NRTKDPNELAEYDVVVTTYAIVTNEAPNKFLVDEDENDEKNTDRYGLASGFSNNKKRK
                VVVGASKKSKRRGRKSTNDTSSEPDCGPLGKVGWFRIVLDEAQTIKNYRTQMARSCCT
                LRAKRRWCLSGTPIQNTIDDLYSYFRFLRYDPYAVYKSFYSTIKVPISRNSCQGYKKL
                QAVLRAIMLRRTKGTLLDGKPIINLPPKVVNLSQVDFSVAERSFYKKLEADSRSQFKA
                YADAGTLSQNYANILLLLLRLRQACDHPQLVKRYNSDPVGKVSEAAVRRLPREARSRL
                INRLESSSAICYECNEPPEKPVVTLCGHIFCYECVLEYITGDENTCPVPRCKQQLARD
                VVFSESSLRNCTSDDSGCSSSHDNGLDRSVFQKRDFCSSKIKAVLDILQSLSQPDSPN
                SAQHGQMPSSSRPYDDDDVTIVEPMRLHSSSPSQGAVKTIIFSQWTGMLDLVELRILE
                SGIEFRRLDGTMSLAARDRAVKEFSKKPDVKVMLMSLKAGNLGLNMVAACHVILLDLW
                WNPTTEDQAIDRAHRIGQTRPVTVTRITIKDTVEDRILKLQEEKRTMVASAFGEEHGG
                SSATRLTVDDLKYLFMV"
BASE COUNT     8810 a   5526 c   6419 g  10767 t
ORIGIN

  Query Match             10.8%;  Score 43.6;  DB 8;  Length 31522;
  Best Local Similarity   76.7%;  Pred. No. 0.092;
  Matches   66;  Conservative    0;  Mismatches   19;  Indels    1;  Gaps    1;

Qy    143 GAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAACCAGATTCCAGGGCCAAG 202
          | || | || || ||| |||||||||||||| || || ||||| || ||   |||||||||
Db  16417 GGAAAGGGAAGTGAAGTTCTTCACAGAGACAGAAGAGAACAATCAAATGAGAGGGCCAAG 16476

Qy    203 GGCTTGGAATCCCATATTTCCTTTGA 228
          | | |||| | || |||||||| |||
Db  16477 GAC-TGGAGTTCCGTATTTCCTATGA 16501


RESULT 10
AC007980/c
LOCUS       AC007980              100975 bp    DNA     linear   PLN 28-FEB-2001
DEFINITION  Arabidopsis thaliana chromosome I BAC F14I3 genomic sequence,
            complete sequence.
ACCESSION   AC007980
VERSION     AC007980.1  GI:5391457
KEYWORDS    HTG.
SOURCE      Arabidopsis thaliana.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 100975)
  AUTHORS   Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
            Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
            Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
            Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vysotskaia,V.S.,
            Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 100975)
  AUTHORS   Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
            Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
            Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
            Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V.,
            Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUL-1999) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
REFERENCE   3  (bases 1 to 100975)
  AUTHORS   Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
            Altafi,H., Nguyen,M., Brooks,S., Buehler,E., Chao,Q., Dunn,P.,
            Gonzalez,A., Khan,S., Kremenetskaia,I., Kim,C., Lenz,C., Li,J.,
            Liu,S., Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V.,
            Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-AUG-1999) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
```

```
            USA
REFERENCE   4  (bases 1 to 100975)
  AUTHORS   Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
            Altafi,H., Nguyen,M., Lam,B., Southwick,A., Ecker,J., Theologis,A.
            and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-FEB-2001) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT     e-mail for correspondence: arab@sequence.stanford.edu
            Genes with similarity to proteins in the databases are described as
            'putative', '-like' or 'similar to'.  Genes that have EST
            similarity but no significant protein similarity are described as
            'unknown proteins'.  Genes that are annotated based only on gene
            prediction software are described as 'hypothetical proteins'.  The
            software programs used to predict genes include:  Grail
            (Informatics Group, Oak Ridge National Laboratory,
            http://compbio.ornl.gov/section/index.html), GENSCAN (Chris Burge,
            http://gnomic.stanford.edu/
            chris/GENSCANW.html), Fexa (V.Solovyev & A.Salamov, Sanger Centre,
            http://genomic.sanger.ac.uk/), and NetPlantGene (S.M. Hebsgaard, et
            al., CBS, Technical University of
            Denmark,http://www.cbs.dtu.dk/NetPlantGene.html).
FEATURES              Location/Qualifiers
     source           1. .100975
                      /organism="Arabidopsis thaliana"
                      /db_xref="taxon:3702"
                      /chromosome="I"
                      /clone="F14I3"
     gene             complement(<1. .4027)
                      /gene="F14I3.1"
     CDS              complement(join(<1. .567,690. .908,1119. .1249,1596. .2526,
                      2612. .2832,2934. .3087,3332. .3876,3937. .4027))
                      /gene="F14I3.1"
                      /note="Similar to transcription factors; Similar to
                      transcription factors.  This BAC contains only the 5-prime
                      portion of the gene"
                      /codon_start=1
                      /protein_id="AAD50036.1"
                      /db_xref="GI:5734771"
                      /translation="MDSAIDISSDSDVEIQETRTRPQHPPRIAEVVDLGAGSHRRDLS
                      TLRPHFLSGSSSGANGHTKTGLTNLDSRNGFESKPLPRAEHHTHIPGNGSIVTSRIPN
                      ISVGDYEKFSSQQAFKRTHPPTFSRPPFPPRPDIGTSNGNASHFRGGAHDDLGMGRVT
                      NGTRILPPSVAHGTSASPSHFNGLSDPMHRNGIGEERNSENDERLIYQAALQTMCMEA
                      LLFFLFDKREGIGTFTIVAFVSYFALLHQGYDFHLKSLVFLHNSPGTESTQVRGRFTC
                      WSSFGSPYEASGIQVYIIFQYLSEWSCVLNTVVSQKIALAWMFQKETNSLHCMGGILA
                      DDQGLGKTVSTIALILKQMHEAKLKSKNSGNQEAEALDLDADDESENAFEKPESKASN
                      GSGVNGDSGIKKAKGEEASTSTRKFNRKRPAAGTLIVCPASVVRQWARELDEKVTDEA
                      KLSVLIYHGGNRTKDPIELAKYDVVMTTYAIVSNEVPKQPLVDDDENDEKNSEKYGLA
                      SGFSINKKRKNVVGTTKKSKKKKGNNNAGDSSDPDSGTLAKVGWFRVVLDEAQTIKNH
                      RTQVARACCGLRAKRRWCLSGTPIQNTIDDLYSYFRFLKYDPYAVYKSFCHQIKGPIS
                      RNSLQGYKKLQAVLRAIMLRRTKGTLLDGQPIINLPPKTINLSQVDFSVEERSFYVKL
                      ESDSRSQFKAYAAAGTLNQNYANILLMLLRLRQACDHPQLVKRYNSDSVGKVSEEAVK
                      KLPKEDLVSLLSRLESSPICCVCHDPPEDPVVTLCGHIFCYQCVSDYITGDEDTCPAP
                      RCREQLAHDVVFSKSTLRSCVADDLGCSSSEDNSHDKSVFQNGEFSSSKIKAVLDILQ
                      SLSNQGTSNSTQNGQMASSSQQPNDDDDDDDDVTIVEKTSLKSTPSNGGPIKTIIFS
                      QWTGMLDLVELSLIENSIEFRRLDGTMSLIARDRAVKEF"
     gene             complement(6533. .8700)
                      /gene="F14I3.2"
     CDS              complement(join(6533. .6586,6685. .6780,6872. .6955,
                      7036. .7104,7261. .7362,7550. .7615,7697. .7747,8020. .8149,
                      8208. .8289,8574. .8700))
                      /gene="F14I3.2"
                      /note="Hypothetical protein"
                      /codon_start=1
                      /protein_id="AAD50049.2"
                      /db_xref="GI:13162531"
                      /translation="MEGFSPPINTAQVDAKTKLDEKVDYSNLPCPVLYEELNREATMA
                      LKPELFEGFRLDYNKSLNQKFFLSHSLCIDASNSILMGPTEVPNPTPSSEIIKIPTAN
                      YDFGAGFIDPKLTDEEDKSQGHLVIDYKGSDYRTQLQLGNNSVYAANYIQHVTPHLSL
                      GGEAFWLGQQLMSGVGYAARYETDKTVASGQIASTGVAVMNYVHKVSEKLSFATDFIY
                      NYLSRDVTASVGYDLITRQSRLRGKVDSNGVVAAYLEEQLPIGLRFLLSAEVDHVKKD
                      YKFGFGVNAF"
     gene             9178. .9869
                      /gene="F14I3.3"
     CDS              join(9178. .9315,9403. .9450,9539. .9655,9732. .9869)
                      /gene="F14I3.3"
                      /note="Similar to fructokinase; Similar to frutokinase,
                      but smaller and lacking some of the sequence of
                      frutokinase"
                      /codon_start=1
                      /protein_id="AAD50037.1"
                      /db_xref="GI:5734772"
                      /translation="MKATEEAKEAGALLSYDPNLREPLWPSPEEARTQIMSIWDKADI
                      IKLLLVTLGEKGCRYYTKDFHGSVETFHVDAVDTTGAGDSFVGALLNQIVDDQSVLEE
                      EERLRKVLRIANACGAITTTKKGAIPALPTDCEALSFLKRQVEQ"
     gene             complement(10670. .14375)
                      /gene="F14I3.4"
     CDS              complement(join(10670. .10748,10841. .10964,11094. .11142,
                      11234. .11299,11516. .11881,11975. .12133,12413. .12704,
                      12911. .13257,13477. .13671,13773. .14056,14204. .14375))
                      /gene="F14I3.4"
                      /note="Similar to oligopeptidases"
                      /codon_start=1
                      /protein_id="AAD50051.1"
                      /db_xref="GI:5734786"
                      /translation="MAESRSPPVAKKVEHVMEMFGDVRVDNYYWLRDDSRTNPDMLSY
                      LREENHYTDFVMSGTKQFENQLFAEIRGRIKEDDISAPLRKGPYYYYEKNLQGKEYIQ
                      HCRRLITDNKAEPSVYDTMPTGPDAPPEHVILDENTKAQEHDYYRIGAFKASPDHKLV
                      AYAEDTKGDEIYTVNVIDSEALKPVGQQLKGLTSYLEWAGNDALLYITMDEILRPDKV
                      WLHKLGTEQSSDVCLYHEKDDMFSLELHASESHKYLFVASESKTTRFVFSLDVSKTQD
                      GLRVLTPRVDGIDSSVSHRGNHFFIQRRSTEFYNSELIACPVDDTSKTTVLLPHRESV
                      KIQEIQLFRDHLAVFERENGLQKITVHRLPAEGQPLEGLQGGRNVSFVDPVYSVDSTE
                      SEFSSRVLRFKYCSMKTPPSVYDYDMDSGTSVVKKIDTVLGGFDASNYVTERKWVAAS
                      DGTQIPMSIVYNKKLAKLDGSDPLLLYGYGSYEISVDPYFKASRLSLLDRGFTFVIAH
                      VRGGGEMGRQWYENGKLLKKKNTFTDFIACAERLIELKYCSKEKLCMEGRSAGGLLMG
                      AVVNMRPDLFKVVIAGVPFVDVLTTMLDPTIPLTTSEWEEWGDPRKEEFYFYMKSYSP
                      VDNVTAQNYPNMLVTAGLNDPRVMYSEPGKWVAKLREMKTDNNVLLFKCELGAGHFSK
                      SGRFEKLQEDAFTFAFMMKVLDMIPASG"
     gene             complement(15183. .17961)
                      /gene="F14I3.5"
     CDS              complement(join(15183. .15266,15621. .15692,15782. .15877,
                      15975. .16118,16224. .16274,16387. .16497,16857. .16914,
                      17131. .17198,17284. .17430,17881. .17961))
                      /gene="F14I3.5"
                      /note="phosphoprotein phosphatase; Location of ESTs
                      VBVRF01, gb|Z33786 and ATTs2101, gb|Z29155.  Protein has
                      greater than 65 % identity with mammalian protein
                      phosphatases"
                      /codon_start=1
                      /protein_id="AAD50050.1"
                      /db_xref="GI:5734785"
                      /translation="MDLDQWISKVKDGQHLSEDELQLLCEYVKEILIEESNVQPVNSP
                      VTVCGDIHGQFHDLMKLFQTGGHVPETNYIFMGDFVDRGYNSLEVFTILLLLKARHPA
                      NITLLRGNHESRQLTQVYGFYDECQRKYGNANAWRYCTDVFDYLTLSAIIDGTVLCVH
                      GGLSPDVRTIDQIRLIERNCEIPHEGPFCDLMWSDPEDIETWAVSPRGAGWLFGSRVT
                      TEFNHINNLDLVCRAHQLVQEGLKYMFQDKGLVTVWSAPNYCYRCGNVASILSFNDNM
                      EREVKFFTETEENNQMRGPRTGVPYFL"
     gene             complement(19749. .26167)
                      /gene="F14I3.6"
     CDS              complement(join(19749. .20405,20494. .20561,20665. .20811,
                      20897. .21026,21113. .21248,21375. .21580,21772. .21978,
                      22125. .22283,22491. .22568,22683. .22772,22907. .23059,
                      23230. .23348,23550. .23688,23778. .23881,24055. .24132,
                      24346. .24409,24517. .24682,24985. .25039,25105. .25255,
                      25382. .25525,25614. .25715,25853. .26167))
                      /gene="F14I3.6"
                      /note="Highly similar to myosin"
                      /codon_start=1
                      /protein_id="AAD50052.1"
                      /db_xref="GI:5734787"
                      /translation="MAHKVKASFQSLKTMPADYRFLGSPISDHLETNLITPPNGHLKN
                      GVNGTASSVGGMDSVNEDSPYSVRSILNGERSSIGDGDSILPLPESNDRKWSDTNVYA
                      RKKVLQFWVQLPNGNWELGKIMSTSGEESVIVVTEGKVLKVKSETLVPANPDILDGVD
                      DLMQLSYLNEPAVLYNLEYRYNQDMIYTKAGPVLVAVNPFKEVPLYGNRNIEAYRKRS
                      NESPHVYAIADTAIREMIRVLSFPVSITDEVNQSIIISGESGAGKTETAKIAMQYLAA
```

```
                     LGGGSGIEYEILKTNPILEAFGNAKTLRNDNSSRFGKLIEIHFSETGKISGAQIQTLV
                     QCTEGERSYHIFYQLCAGASPTLREKLNLTSAKQYNYLKQSNCYSINGVDDAERFHAV
                     KEALDIVHVSKEDQENVFAMLAAVLWLGNVSFTIIDNENHVEPEPDESLSTVAKLIGC
                     NINELKLALSKRNMRVNNDTIVQKLTLSQAIDARDALAKSIYACLFDWLVEQINKSLA
                     VGKRRTGRSISILDIYGFESFNKNSFEQFCINYANERLQQHFNRHLFKLEQEEYIQDG
                     IDWTRVDFEDNQECLSLFEKKPLGLLSLLDEESTFPNGTDLTLANKLKQHLNDNSCFR
                     GDRGKAFTVAHYAGEVTYETTGFLEKNRDLLHSDSIQLLSSCSCHLPQAFASSMLIYS
                     EKPLVGPLHKAGGADSQRLSVATKFKGQLFQLMQRLGNTTPHFIRCIKPNNVQSAGLY
                     EQGLVLQQLRCCGVLEVVRISRSGFPTRMFHHKFARRYGFLLLENIAAKDPLSVSVAI
                     LHQFNILPEMYQVGYTKLFFRTGQIGVLEDTRNRTLHGILRLQSYFRGHQARCRLKEL
                     KTGITILQSFVRGEKMRKEYTELLQRHRASAAIQSHVKRRIASQQYKATVDASAVIQS
                     AIRGELVRRCAGDIGWLSSGGTKRNESDEVLVKASYLSDLQRRVLRTEAALREKEEEN
                     DILRQRVQQYDNRWSEYETKMKSMEEIWQKQMKSLQSSLSIAKKSLEVEDSARNSDAS
                     VNASDATDLDSGGSHYQMGHGRSRSVGVGLSVISRLAEEFGQRAQVFGDDRKFLMEVK
                     SGQVEANLNPDRELRRLKQMFETWKKDYGGRLRETKLILSKLGSEETGGSAEKVKMNW
                     WGRLRSTRY"
     gene            30400. .32418

  Query Match           10.6%;  Score 42.8;  DB 8;  Length 100975;
  Best Local Similarity  68.6%;  Pred. No. 0.17;
  Matches   59;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy      63 TTGAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATAT 122
           ||| |   |||||||||||| ||  | |  |    | | ||||| ||||  ||||||||||
Db   15700 TTGTAAAGGTGTGGTCTGCGCCTAATTACTGCTACCGTTGTGGGAATGTGGCTTCTATAT 15641

Qy     123 TAAGCTTCGATGAGAAGAAGGAAAGG 148
           | || ||| |||| || | ||  |||
Db   15640 TGAGTTTCAATGACAACATGGTGAGG 15615


RESULT 11
AP003740/c
LOCUS       AP003740              108992 bp    DNA     linear   PLN 03-AUG-2002
DEFINITION  Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 7,
            BAC clone:OJ1118_B03.
ACCESSION   AP003740
VERSION     AP003740.2  GI:21202839
KEYWORDS    .
SOURCE      Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA,
            clone:OJ1118_B03.
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC
            clone:OJ1118_B03
  JOURNAL   Published Only in Database (2001)
REFERENCE   2  (bases 1 to 108992)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUN-2001) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     On May 24, 2002 this sequence version replaced gi:14422458.
            Genes were predicted from the integrated results of the following:
            GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH
            (http://www.softberry.com/), GeneMark.hmm
            (http://opal.biology.gatech.edu/GeneMark/), GlimmerM
            (http://www.tigr.org/tdb/glimmerm/glmr_form.html), RiceHMM
            (http://rgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor
            (http://bioinformatics.iastate.edu/cgi-bin/sp.cgi), BLASTN and
            BLASTX. The genomic sequence was searched against NCBI NonRedundant
            Protein database, nr
            (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
            RGP. Protein homologies of the coding regions were searched against
            NCBI NonRedundant Protein database with BLASTP. ESTs represent the
            identified cDNA sequences using BLASTN with the corresponding DDBJ
            accession no. and RGP clone ID.
            A gene with identity or significant homology to a protein is
            classified based on the protein name to indicate the homology level
            such as same name, 'putative-' and '-like protein'.  A gene without
            significant homology to any protein but with EST homology (covering
            almost the entire length of partial sequence) is classified as an
            'unknown' protein. A gene predicted by two or more gene prediction
            programs is classified as a 'hypothetical' protein according to
            IRGSP standard. A gene predicted by a single gene prediction
            program is also classified as a probable 'hypothetical' protein and
            is included as a miscellaneous feature of the sequence.
            The orientation of the sequence is from -21M13 to M13rev of the BAC
            clone. This sequence of OJ1118_B03 clone has an overlap with
            P0592C06 (DDBJ:AP005197) clone at the 5' end and with P0668C05
            (DDBJ:AP004572) at the 3' end. The sequence was generated by
            combining Monsanto and RGP-Japan sequencing data. Detailed
            information on overlap and assembly quality together with
            annotation of this entry is available at
            http://rgp.dna.affrc.go.jp/GenomeSeq.html.
FEATURES            Location/Qualifiers
     source         1. .108992
                    /organism="Oryza sativa (japonica cultivar-group)"
                    /cultivar="Nipponbare"
                    /db_xref="taxon:39947"
                    /chromosome="7"
                    /clone="OJ1118_B03"
     LTR            complement(3276. .3488)
                    /note="3' LTR"
     gene           complement(3946. .8619)
                    /gene="OJ1118_B03.1"
     misc_feature   complement(3946. .8619)
                    /gene="OJ1118_B03.1"
                    /note="probably inactive due to frameshift(s) in CDS
                    probably inactive due to stop codon(s) in CDS
                    pseudogene,copia polyprotein"
                    /pseudo
     gene           join(9485. .9573,9643. .9940,10027. .10092,10185. .10196)
                    /gene="OJ1118_B03.2"
     CDS            join(9485. .9573,9643. .9940,10027. .10092,10185. .10196)
                    /gene="OJ1118_B03.2"
                    /note="hypothetical protein
                    similar to Oryza sativa chromosome1,BAA90382"
                    /codon_start=1
                    /protein_id="BAC06913.1"
                    /db_xref="GI:22093617"
                    /translation="MRARRRWLVVTAGSTELVSAAGSTIEAERRSLLCAGGVWFGMRA
                    GEAESYRQREATAASTPIPVACGSAGGVPCVPAAGDADAVSSASLRAAAVNGECRRSI
                    EDPSAGFSRANRRWKADGTILRPSVKGDDGGTSMRIWADSFGIGDHVSQRLV"
     LTR            complement(10633. .10839)
                    /note="5' LTR"
     gene           join(11450. .11597,12715. .12785,12991. .13110,13298. .13416,
                    13632. .13713,13847. .13852)
                    /gene="OJ1118_B03.3"
     CDS            join(11450. .11597,12715. .12785,12991. .13110,13298. .13416,
                    13632. .13713,13847. .13852)
                    /gene="OJ1118_B03.3"
                    /note="contains ESTs D40795(S2947),AU081605(S2947)"
                    /codon_start=1
                    /product="ADP-ribosylation factor 1"
                    /protein_id="BAC06914.1"
                    /db_xref="GI:22093618"
                    /translation="MGLAFGKLFSRLFAKKEMRILMVGLDAAGKTTILYKLKLGEIVT
                    TIPTIGFNVETVEYKNISFTVWDVGGQDKIRPLWRHYFQNTQGLIFVVDSNDRDRVVE
                    ARDELHRMLNEDELRDAVLLVFANKQDLPNAMNAAEITDKLGLHSLRQRHWYIQSTCA
                    TTGEGLYEGLDWLSSNIASKA"
     gene           join(15518. .15613,15722. .15796)
                    /gene="OJ1118_B03.4"
     CDS            join(15518. .15613,15722. .15796)
                    /gene="OJ1118_B03.4"
                    /note="hypothetical protein
                    predicted by GeneMark.hmm etc."
                    /codon_start=1
                    /protein_id="BAC06915.1"
                    /db_xref="GI:22093619"
```

```
                /translation="MGVVKSPAVVFLVESDQLKKIYPEESEFNEDAESFFLVLQGLVS
                PYGYKYVYPGSF"
     gene       complement(join(16653. .16790,16994. .17128))
                /gene="OJ1118_B03.5"
     CDS        complement(join(16653. .16790,16994. .17128))
                /gene="OJ1118_B03.5"
                /note="hypothetical protein
                predicted by GeneMark.hmm etc."
                /codon_start=1
                /protein_id="BAC06916.1"
                /db_xref="GI:22093620"
                /translation="MKMTQGDIHLQAAGALDGEASIAGPATTGECAWMATRSCAALTA
                LVQRLVRCGRASQFDPATDKIYKQQINKPIGCQADGAIPADADTSR"
     gene       19388. .19714
                /gene="OJ1118_B03.6"
     CDS        19388. .19714
                /gene="OJ1118_B03.6"
                /note="hypothetical protein
                similar to Oryza sativa chromosome 1,B1156H12.26"
                /codon_start=1
                /protein_id="BAC06917.1"
                /db_xref="GI:22093621"
                /translation="MDAGSASCRRSLLGPCAACVRPRRAWPRLTRAGRQEGEEAALAG
                HRMGREEGEERAGMGGGGGGEGDGSEERRGRLLRRLRPVVAERIGNKDYILCVILVTV
                GNTNRE"
     gene       22328. .23398
                /gene="OJ1118_B03.7"
     CDS        22328. .23398
                /gene="OJ1118_B03.7"
                /note="hypothetical protein
                predicted by GlimmerM etc."
                /codon_start=1
                /protein_id="BAC06918.1"
                /db_xref="GI:22093622"
                /translation="MASSSSGKRVSSSSTSSHQQHAKRRRRAPGDELSQTHGVARNAV
                AAMPMVGVAPLMLSPVTMQAQLLCYVAMPMAPVAGGGVPLAAALCRLRNGTLERMMRE
                MAQECTPPLVARRDRKSGATPPPPPPWWPTAEEPWWGTEVAAHLRGRRMATPVPFASP
                RRLAKAEKVAVLVAVVRHVAPDFGRLAAAAGRSRLTELESSSIWESALRGERERRYVVM
                PSFILLPPPPPPPPQHAHSAESAAAHAGPEPESADQTAVDFSVSGEDTVTTGPELEQL
                VGDGNGATVEGEGQKTEDSPVPPLLEQHGGEDHDHHRQHGGFSGEVSGAAPEDVDWFD
                YDEVLRGLDELEIPSFFGGYYI"
     gene       join(25014. .25018,25374. .25602)
                /gene="OJ1118_B03.8"
     misc_feature join(25014. .25018,25374. .25602)
                /gene="OJ1118_B03.8"
                /note="This category is not included in IRGSP standard.
                hypothetical ORF
                predicted by GENSCAN"
     gene       join(29910. .30038,31430. .31621,32179. .32262,32296. .32496,
                32801. .32899,33193. .33231)
                /gene="OJ1118_B03.9"
     CDS        join(29910. .30038,31430. .31621,32179. .32262,32296. .32496,
                32801. .32899,33193. .33231)
                /gene="OJ1118_B03.9"
                /note="hypothetical protein
                predicted by FGENESH etc."
                /codon_start=1
                /protein_id="BAC06919.1"
                /db_xref="GI:22093623"
                /translation="MRSCCCCCPCPAALGIGIGPRLRSFLRDYDALQSLALALIYLQI
                GCALIGSLGALFNGVLVINLVIGLFAVVAIESSSQTLGRTYAVLLFFAIVLDVAWFIL
                FSHAINITPEEKYGQLFVLSLKLALWMQIIGFSMYRLGVSSSTPTYHEVNYDGRNSFL
                SPRSSSVRRNSMADDILGGSIYDPAYYSSLFADVRNNTCTHQGDKQSGSNDSGSTSAG
                QSPRLKSFASRSFVANDLEDHTFQFSDFS"
     gene       join(36427. .36609,38877. .38975,39825. .39950,40746. .40820,
                40868. .41155,41199. .41801)
                /gene="OJ1118_B03.10"
     CDS        join(36427. .36609,38877. .38975,39825. .39950,40746. .40820,
                40868. .41155,41199. .41801)
                /gene="OJ1118_B03.10"
                /note="hypothetical protein
                predicted by FGENESH etc."
                /codon_start=1
                /protein_id="BAC06920.1"
                /db_xref="GI:22093624"
                /translation="MSAITTELLVLDDAAAGASKAAPGTGEARAAPDPRHHRRPPRST
                SSSPPREEGPDASAGGTSFTLKTIQIYFQRPKVQAFMVWPYDPRAILMLEKRQLPVDR
                VDSHAGREEELLPRRIAARRALADDTDSINQYDQQRGQFFLHKGLMKREIFYKTQHME
                REKSKQQKIEKRAKTCIPYCLLPPNTRTSQFTCRPIEESQEKKQEIGKEHCAEPRNRL
                RQSAGIELYLTKSQELSHYKESRNADMPRTQIRLGGRRRRGSAARWFATGGRSRGGRR
                GYIRGGRALAVRPSGHHDVDADRAELRLILLGIPDHQRANPTHHRRLLLRSANTLERN
                TLKKKPKEAVTTEEKVTEPGEKQLSPNPPNRVTSNRESGTTKGKKRLPGGEQHTPTPK
                VQQQRMKTAHPQDGLTSTDSGTEDDKQQRPPARASHQRRGDNTNPKRTRKTQKSQQGK
                STCTYPA"
     gene       join(45022. .45585,48876. .48930,51913. .51950)
                /gene="OJ1118_B03.11"
     CDS        join(45022. .45585,48876. .48930,51913. .51950)

  Query Match          10.5%;  Score 42.4;  DB 8;  Length 108992;
  Best Local Similarity   72.4%;  Pred. No. 0.22;
  Matches   55;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy    166 CAGAGACCGAGGAAAACAACCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTT 225
          || |||| ||||| ||  ||||||| | ||| |||| |  |     |||| ||||||||
Db 101614 CATAGACAGAGGAGAATGACCAGATGCAAGGCCCAACGATTGTTTGTCCCGTATTTCCTC 101555

Qy    226 TGACTTCAGGTGGACA 241
          ||||||||  || |||
Db 101554 TGACTTCAATTGTACA 101539


RESULT 12
AP004572/c
LOCUS       AP004572              142081 bp    DNA     linear   PLN 17-AUG-2002
DEFINITION  Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 7,
            PAC clone:P0668C05.
ACCESSION   AP004572
VERSION     AP004572.3  GI:22296450
KEYWORDS    .
SOURCE      Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA,
            clone:P0668C05.
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, PAC
            clone:P0668C05
  JOURNAL   Published Only in Database (2001)
REFERENCE   2  (bases 1 to 142081)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-DEC-2001) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     On Aug 16, 2002 this sequence version replaced gi:21623780.
            Genes were predicted from the integrated results of the following:
            GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH
            (http://www.softberry.com/), GeneMark.hmm
            (http://opal.biology.gatech.edu/GeneMark/), GlimmerM
            (http://www.tigr.org/tdb/glimmerm/glmr_form.html), RiceHMM
            (http://rgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor
            (http://bioinformatics.iastate.edu/cgi-bin/sp.cgi), BLASTN and
            BLASTX. The genomic sequence was searched against NCBI NonRedundant
            Protein database, nr
            (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
            RGP. Protein homologies of the coding regions were searched against
            NCBI NonRedundant Protein database with BLASTP. ESTs represent the
            identified cDNA sequences using BLASTN with the corresponding DDBJ
            accession no. and RGP clone ID.
            A gene with identity or significant homology to a protein is
            classified based on the protein name to indicate the homology level
            such as same name, 'putative-' and '-like protein'.  A gene without
```

```
            significant homology to any protein but with EST homology (covering
            almost the entire length of partial sequence) is classified as an
            'unknown' protein. A gene predicted by two or more gene prediction
            programs is classified as a 'hypothetical' protein according to
            IRGSP standard. A gene predicted by a single gene prediction
            program is also classified as a probable 'hypothetical' protein and
            is included as a miscellaneous feature of the sequence.
            The orientation of the sequence is from T7 to SP6 of the PAC clone.
            This sequence of P0668C05 clone has an overlap with OJ1118_B03
            (DDBJ:AP003740) clone at the 5' end. Detailed information on
            overlap and assembly quality together with annotation of this entry
            is available at
            http://rgp.dna.affrc.go.jp/GenomeSeq.html.
FEATURES                Location/Qualifiers
     source             1. .142081
                        /organism="Oryza sativa (japonica cultivar-group)"
                        /cultivar="Nipponbare"
                        /db_xref="taxon:39947"
                        /chromosome="7"
                        /clone="P0668C05"
     gene               join(3987. .4169,6437. .6535,7385. .7510,8306. .8380,
                        8428. .8715,8759. .9361)
                        /gene="P0668C05.1"
     CDS                join(3987. .4169,6437. .6535,7385. .7510,8306. .8380,
                        8428. .8715,8759. .9361)
                        /gene="P0668C05.1"
                        /note="hypothetical protein
                        predicted by FGENESH etc."
                        /codon_start=1
                        /protein_id="BAC10217.1"
                        /db_xref="GI:22296451"
                        /translation="MSAITTELLVLDDAAAGASKAAPGTGEARAAPDPRHHRRPPRST
                        SSSPPREEGPDASAGGTSFTLKTIQIYFQRPKVQAFMVWPYDPRAILMLEKRQLPVDR
                        VDSHAGREEELLPRRIAARRALADDTDSINQYDQQRGQFFLHKGLMKREIFYKTQHME
                        REKSKQQKIEKRAKTCIPYCLLPPNTRTSQFTCRPIEESQEKKQEIGKEHCAEPRNRL
                        RQSAGIELYLTKSQELSHYKESRNADMPRTQIRLGGRRRRGSAARWFATGGRSRGGRR
                        GYIRGGRALAVRPSGHHDVDADRAELRLILLGIPDHQRANPTHHRRLLLRSANTLERN
                        TLKKKPKEAVTTEEKVTEPGEKQLSPNPPNRVTSNRESGTTKGKKRLPGGEQHTPTPK
                        VQQQRMKTAHPQDGLTSTDSGTEDDKQQRPPARASHQRRGDNTNPKRTRKTQKSQQGK
                        STCTYPA"
     gene               join(12582. .13145,16436. .16490,19473. .19510)
                        /gene="P0668C05.2"
     CDS                join(12582. .13145,16436. .16490,19473. .19510)
                        /gene="P0668C05.2"
                        /note="hypothetical protein
                        predicted by GlimmerM etc."
                        /codon_start=1
                        /protein_id="BAC10218.1"
                        /db_xref="GI:22296452"
                        /translation="MMMDHNILIPLVSSLMMLVLGPLIIDVISVSNKIGRLFSAIARF
                        LAHDDSVINSMVVDNDPPPSSQLLGGGVSCFDAMTVTTRLGLRWQRSGEAAMECQGCD
                        IPMDATVDELLDRKMASEGELKDAFYVFDRNEDGFICASELWSVMRRLGFKEGQRYED
                        CMRMIHTFDEDRDGRISYLEFRRMMEDADGCALYFGIEGVLVILSTGVCGERAAVGGV
                        "
     gene               join(21422. .21505,21590. .21691,21971. .22119,22452. .22599)
                        /gene="P0668C05.3"
     CDS                join(21422. .21505,21590. .21691,21971. .22119,22452. .22599)
                        /gene="P0668C05.3"
                        /note="contains ESTs D23260(C2515),C98364(C2515)"
                        /codon_start=1
                        /product="putative 60S ribosomal protein L24"
                        /protein_id="BAC10219.1"
                        /db_xref="GI:22296453"
                        /translation="MYRTELCRFSGAKIYPGKGIRFIRADSQVFLFSNSKCKRYFHNR
                        LKPAKLTWTAMYRKQHKKDIHAEAVKKRRRTTKKPYSRSIVGATLEVIQKKRAEKPEV
                        RDAARESALREIKERIKKTKDEKKAKKAEVAKSQKASGKGNAPKPGKGPKLGGGGGKR
                        "
     gene               join(24628. .24657,25452. .25481,26190. .26322,27451. .27491)
                        /gene="P0668C05.4"
     misc_feature       join(24628. .24657,25452. .25481,26190. .26322,27451. .27491)
                        /gene="P0668C05.4"
                        /note="This category is not included in IRGSP standard.
                        hypothetical ORF
                        predicted by GENSCAN"
     gene               join(28528. .28547,28852. .28939,29391. .29495)
                        /gene="P0668C05.5"
     misc_feature       join(28528. .28547,28852. .28939,29391. .29495)
                        /gene="P0668C05.5"
                        /note="This category is not included in IRGSP standard.
                        hypothetical ORF
                        predicted by GENSCAN"
     gene               complement(join(30351. .30405,31104. .31280,31382. .31464))
                        /gene="P0668C05.6"
     misc_feature       complement(join(30351. .30405,31104. .31280,31382. .31464))
                        /gene="P0668C05.6"
                        /note="This category is not included in IRGSP standard.
                        hypothetical ORF
                        predicted by GlimmerM"
     LTR                complement(40109. .40452)
                        /note="3' LTR"
     gene               complement(40452. .44645)
                        /gene="P0668C05.7"
     misc_feature       complement(40452. .44645)
                        /gene="P0668C05.7"
                        /note="probably inactive due to frameshift(s) in CDS
                        probably inactive due to stop codon(s) in CDS
                        pseudogene, gypsy/Ty-3 retroelement polyprotein"
                        /pseudo
     LTR                complement(45109. .45453)
                        /note="5' LTR"
     gene               50849. .51169
                        /gene="P0668C05.8"
     CDS                50849. .51169
                        /gene="P0668C05.8"
                        /note="hypothetical protein
                        predicted by GlimmerM etc."
                        /codon_start=1
                        /protein_id="BAC10220.1"
                        /db_xref="GI:22296454"
                        /translation="MDPAAARPSAPGCRRSGTGPAAIAPHGPGRCRPSPAQARPPPPP
                        PGMAPPPLPLLGTGPAAAAPAQHGPGRCLGCFSRRSEEEKRGKLDQMWKRKGRGRDES
                        ERGR"
     gene               complement(join(55139. .55378,60225. .60296))
                        /gene="P0668C05.9"
     CDS                complement(join(55139. .55378,60225. .60296))
                        /gene="P0668C05.9"
                        /note="hypothetical protein
                        predicted by GeneMark.hmm etc."
                        /codon_start=1
                        /protein_id="BAC10221.1"
                        /db_xref="GI:22296455"
                        /translation="MEGTRRHLEKDLEPWRIIEDVQVKRVETEAMATTLEEEGRGVGA
                        WRALEPEAEAAEGKGGGGADAGGGIAVGVNSTGRRGWGWEGLGGWEDGAGGWLDKVVG
                        G"
     gene               complement(join(63541. .63724,63791. .63849))
                        /gene="P0668C05.10"
     CDS                complement(join(63541. .63724,63791. .63849))
                        /gene="P0668C05.10"
                        /note="hypothetical ORF"
                        /codon_start=1
                        /protein_id="BAC10222.1"
                        /db_xref="GI:22296456"
                        /translation="MSIVLPILHNRKELEFLNVRDASVGMLHMSIPHLATAARFDHRS
                        VIGQSKIVLLPTNLNQIAISSMEGGLFDTFPANSEI"
     gene               join(66254. .66289,67073. .67201,67242. .67324,68430. .68550,
                        68861. .69007)
                        /gene="P0668C05.11"
     CDS                join(66254. .66289,67073. .67201,67242. .67324,68430. .68550,
                        68861. .69007)
                        /gene="P0668C05.11"
                        /note="hypothetical ORF"
                        /codon_start=1
                        /protein_id="BAC10223.1"
                        /db_xref="GI:22296457"
                        /translation="MGIPASLLTSAQINPNQPSQRVRRRRLAGATGSCSGGDSDVLLR
                        TSAVTGGRYGRARTAAASGGRVRPRQAAAFYGSDDQQVMLWMDRSKVRSSMTSSSVLS
```

```
                EMSVSCVVLGGCAILSNVEMLLNLLLVSPLISFSLAVSLRTYTGGGANSIRQLTFASN
                HPSNIYRILFK"
     gene       70724. .70885
                /gene="P0668C05.12"
     CDS        70724. .70885
                /gene="P0668C05.12"
                /note="hypothetical ORF"
                /codon_start=1
                /protein_id="BAC10224.1"
                /db_xref="GI:22296458"
                /translation="MADPLAGRWHPLEKRRGEGRRRDWKRGEGEEELARRSGKRRERI

  Query Match          10.5%;  Score 42.4;  DB 8;  Length 142081;
  Best Local Similarity  72.4%;  Pred. No. 0.23;
  Matches   55;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy    166 CAGAGACCGAGGAAAACAACCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTT 225
          || |||| ||||| ||  ||||||| | ||| |||| |  |     |||| ||||||||
Db  69174 CATAGACAGAGGAGAATGACCAGATGCAAGGCCCAACGATTGTTTGTCCCGTATTTCCTC 69115

Qy    226 TGACTTCAGGTGGACA 241
          ||||||||  || |||
Db  69114 TGACTTCAATTGTACA 69099


RESULT 13
AC091724/c
LOCUS       AC091724              158294 bp    DNA     linear   PLN 06-FEB-2002
DEFINITION  Oryza sativa chromosome 10 clone OSJNBa0004E08, complete sequence.
ACCESSION   AC091724
VERSION     AC091724.3  GI:18464007
KEYWORDS    HTG.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 158294)
  AUTHORS   Wing,R.A., Yu,Y., Soderlund,C., Chen,M., Kim,H.-R., Rambo,T.,
            Saski,C., Henry,D., Oates,R., Simmons,J., Wilson,R., Minx,P. and
            Kyung,K.
  TITLE     Rice Genomic Sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 158294)
  AUTHORS   Wing,R.A., Frisch,D., Presting,G., Wood,T., Yu,Y., Soderlund,C.,
            Kim,H.-R., Rambo,T., Henry,D. and Simmons,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAY-2001) Clemson University Genomics Institute,
            Clemson University, 100 Jordan Hall, Clemson, SC 29634, USA
REFERENCE   3  (bases 1 to 158294)
  AUTHORS   Wing,R.A., Yu,Y., Soderlund,C., Chen,M., Kim,H.-R., Rambo,T.,
            Saski,C., Henry,D., Oates,R., Simmons,J., Wilson,R., Minx,P. and
            Kyung,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JAN-2002) Clemson University Genomics Institute,
            Clemson University, 100 Jordan Hall, Clemson, SC 29634, USA
REFERENCE   4  (bases 1 to 158294)
  AUTHORS   Wing,R.A., Yu,Y., Soderlund,C., Chen,M., Kim,H.-R., Rambo,T.,
            Saski,C., Henry,D., Oates,R., Simmons,J., Wilson,R., Minx,P. and
            Kyung,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-FEB-2002) Clemson University Genomics Institute,
            Clemson University, 100 Jordan Hall, Clemson, SC 29634, USA
REFERENCE   5  (bases 1 to 158294)
  AUTHORS   Wing,R.A., Yu,Y., Soderlund,C., Chen,M., Kim,H.-R., Rambo,T.,
            Saski,C., Henry,D., Oates,R., Simmons,J., Wilson,R., Minx,P. and
            Kyung,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-FEB-2002) Clemson University Genomics Institute,
            Clemson University, 100 Jordan Hall, Clemson, SC 29634, USA
REFERENCE   6  (bases 1 to 158294)
  AUTHORS   Wing,R.A., Yu,Y., Soderlund,C., Chen,M., Kim,H., Rambo,T.,
            Saski,C., Henry,D., Oates,R., Simmons,J., Wilson,R., Minx,P.,
            Kyung,K., Thurmond,S.K. and Sun,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-2002) Clemson University Genomics Institute,
            Clemson University, 100 Jordan Hall, Clemson, SC 29634, USA
COMMENT     On Feb 1, 2002 this sequence version replaced gi:18376840.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double -stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality
            >=30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by more than
            one plasmid subclone; and the assembly was confirmed by restriction
            digest.  The following regions are single stranded and below phred
            30: 120374-120415, 118002 -118003 and 22860-22861:  Base 22860 is
            tagged ambiguous.
FEATURES             Location/Qualifiers
     source          1. .158294
                     /organism="Oryza sativa"
                     /db_xref="taxon:4530"
                     /chromosome="10"
                     /clone="OSJNBa0004E08"
     gene            8731. .9276
                     /gene="OSJNBa0004E08.1"
                     /note="Contains similarity to cyclophilin"
     CDS             8731. .9276
                     /gene="OSJNBa0004E08.1"
                     /codon_start=1
                     /product="Putative cyclophilin"
                     /protein_id="AAL75728.1"
                     /db_xref="GI:18542886"
                     /translation="MAPAASSKSNPRVFLDIAIGGEWVGRVVIELLADKVPDTAENFR
                     RLCTGERAGRSGKSRLHYKGSAFHRVVPGFMCQGGDITAGNGTGGESALDGAARHFPD
                     EGFAVKHDGPGVVSMANAGPNTNGSQFFITVDKAPWLDGRHVAFGRVVAGMGAVRAID
                     RTGTWSGKTVKPVVITDCGVL"
     gene            16649. .21066
                     /gene="OSJNBa0004E08.2"
                     /note="Contains similarity to TNP2-like transposable
                     element"
     CDS             join(16649. .16852,17861. .18869,19010. .21066)
                     /gene="OSJNBa0004E08.2"
                     /codon_start=1
                     /product="Putative TNP2-like transposable element"
                     /protein_id="AAL75729.1"
                     /db_xref="GI:18542887"
                     /translation="MVNAEPNTNGSQFIIVNRPPWLDGRHVVFRHVVDGMDIVRAVER
                     TATWRGKMVKPVVIGLCGKALALLMDRQWMYADRRSKEFIHGVHYFLRVAEANRQRGF
                     ICCPCNKCKNQKEYSASRTIHFHLFESGFMPSYNCWTSHGEQGVEIEEDEVEDDNIPD
                     FAQYVGFEGNQTGEEEIAADGNDVADDLGQMLQDAREDCESEKEAHKLDKMLEDHRTS
                     LYPGCEQGHKKLDTTLELLQWKAKNGVSDKAFGDLLKLVKNILPGGNKLPETTYEAKK
                     IVCPLGLEVHKIHACPNDCILYRGEEYENLEACPVCKALRYKIRRDDPGEVDGQLTKK
                     RIPAKVMWYFPIIPRLRRLFRNKGNAGMLRWHAEERQQDGMLRHPADGSQWRNIDRKF
                     KEFGKDARNIRFGPKQPGNDIDVYLRPLVEDLKQLWKKEGVPVWDEDKQEEFNLRALL
                     FVTINDWPALSNLSGQSNKGYKACTHCMDETESTYLKHCRKVVYMGHRRFLAANLPVR
                     KKGKHFEHKADHHTKPKHRSGKTVFAMVKDLKVVFGKGPGSQHIESEDGHAAMWKKNS
                     IFWELPYWEFLDVRHAIDVMHLTKNLCVNLLGFLGVYGKSKDTLEARNDLKHMEQRGD
                     LHPEPKEKGSHYLSPASYTLSKAEKESMFECLESIKVPSGYSTNIKLIISMKEKKFTN
                     LKSHDCHVLMTQLLPVVIRGILPDNVRATITKLCAFMNAISQKVIDPDRLEALQNEVV
                     QCLVSFELIFPPSFFNIMTHLLCHLVKEIRILGPMYLHNMFPFERYMGVLKKYVRNRA
                     RPEASIAKGYGTEEVIEFCVEFIEDLRPIGVPESRHEGRLRGKGTLGRKAIMTVDNNL
                     FRKAHFTVLQHSSLVAPYIEEHLALVRARNIGKSDAWITRHHIDTFPAWLRQHLMGNE
                     SINQQLAFLARGPSGSIATFQGYEINGYTFYTRAQDMKSTNQNSAVRVDAMGHDGTTA
                     TYYGAIEDIWELDYGPLKVPLFRCQWVRLTGGGVMIDDSGMTTVVLNKVGYSDEPFVL
                     ANDVTQVFFVKDMSRKGKKGRGPDEPKRQVVLPTKEKSSELRTRLTRITISWMGNPLL
                     W"
     repeat_region   16649. .21066
                     /note="Contains similarity to TNP2-like transposable
                     element"
     gene            22772. .23155
                     /gene="OSJNBa0004E08.3"
                     /note="Unknown protein"
     CDS             22772. .23155
                     /gene="OSJNBa0004E08.3"
                     /codon_start=1
                     /product="Unknown protein"
```

```
                    /protein_id="AAL75730.1"
                    /db_xref="GI:18542888"
                    /translation="MPPPPRHAAAATPRHAAAATAPQRHAAAAVATPPPPPPPPRRRA
                    NRRPDAATPPLTNENENDRTNESGNEHVNVRCRENVNENENDRTNESENENDRSKTSE
                    NEGTNVNENKRTNVNENERTNEDER"
repeat_region       22866. .22889
                    /note="Simple Sequence Repeat (CGC)n"
repeat_region       23167. .23250
                    /note="Simple Sequence Repeat (AT)n"
gene                24059. .27759
                    /gene="OSJNBa0004E08.4"
                    /note="Contains similarity to TNP2-like transposable
                    element"
CDS                 join(24059. .25480,25555. .26654,26764. .26937,27034. .27144,
                    27229. .27301,27396. .27494,27574. .27759)
                    /gene="OSJNBa0004E08.4"
                    /codon_start=1
                    /product="Putative TNP2-like transposable element"
                    /protein_id="AAL75731.1"
                    /db_xref="GI:18542889"
                    /translation="MADRDEEQILYDTIAEGSSQYWNEEEGNEDPNQYLNEEGNVERD
                    AEGNQQGHVERDVEGNQEEEASGSQPSVGQKRARGQRGAAKKLEGRHIITEVQEDGRP
                    SAPAEAAKNYVRHSGWVVRDNVPVSRVYWRRTRARRDHESFVPDSEKEMMWTTMLETF
                    TLPAGTEDKVKRWTLKKMAEQFQSFKGDMYQKYILKGQTPNFDTFPKLRDHWDEFIAY
                    KTGEQGQAMMERNKENAAKKKYHHHLGSGGYSVAMPKWEQMEASLIERGIEPATANWP
                    ERSKFWYYAHGGTLNPADGSLVFGDQIREAARRLTDAVEASSQGTFRPDRDRDELTLA
                    LQTPEHPGRTRGKGVIPWKIGFKEDIHTYRSRMRSKRDTEAKIADLEFRVSSYELNMQ
                    EEVARKVDERMAAHRSHDPQPTIPPAMVSPSGNRSSCASTGQVGSQSMDAMQTQDEST
                    CPVDDITQRTPCELHIPFKNLSIKVASGMAIPTDPSGTYHCRPIPAGYSKVEVELVEG
                    AYEDLELDYPGGDGETHLRDTSHAIILWRRRYIILPGRQAASRAPSPLAPPSPPQDPA
                    PSPPHAPPAPSPPQAPASTPPLDPAPTPPRAPTPTPPQAPLPAPSKSRAPPAPPPAHT
                    RATKKAKVDAAKNKDPGYDCTQEELDAYVASEVKRQFKPRSPEKKIPIDPSVRNFFRG
                    MSASVKEAIKLSDYERTLKKASSGKSKPVPQLGEQPNQEIEPLVTGKEMTIEQFITDT
                    GLTTDQLLGVAPIEKAEVKYMYELGKPLVKPELLQSLPTQMYKFHQLYMEMSATGREM
                    IGARIRDTDFLQGDDILWINFRGIYELYQLDALDVSIMSCWILMEIQRARRRRVFDTG
                    FIDPRKVNVAMLDQYPQETEDNLVHLLKAQHYKTFILLRYNTEFHWVLLLFDLEACTV
                    NVYDSMDKKESTFDKVFELIDRAWYRFRHLVRGKWRERLRRKFKFPCTKQKQGTNLCG
                    YYVCEYCHCLADQIITTRELDFIRMRDNLTTHKEFIAAVQEQLMGFINEEILDPKGEF
                    YYDGNTIHRSLASELAASTTTSKS"
repeat_region       24059. .27759
                    /note="Contains similarity to TNP2-like transposable
                    element"
repeat_region       27948. .27967
                    /note="Simple Sequence Repeat (AT)n"
repeat_region       28020. .28065
                    /note="Simple Sequence Repeat (TA)n"
gene                29297. .30535
                    /gene="OSJNBa0004E08.5"
                    /note="Hypothetical protein"
CDS                 join(29297. .29344,29819. .30102,30427. .30535)
                    /gene="OSJNBa0004E08.5"
                    /codon_start=1
                    /product="Hypothetical protein"
                    /protein_id="AAL75732.1"
                    /db_xref="GI:18542890"
                    /translation="MARWPAFSSSSAAASTARLRARRGGDDHICGYAEGSLSLVAHPT
                    GLAAVACAATRLSAAGTKSAVRVMHAMARHSATPAVLQEMLAVGVGARLLFLVQVGAS
                    GEQTRARASNVLARWDLGTGGNNFLGHEFTGLYRTRSPGQFVRK"
gene                33040. .34164
                    /gene="OSJNBa0004E08.6"
                    /note="Unknown protein"
CDS                 33040. .34164
                    /gene="OSJNBa0004E08.6"
                    /codon_start=1
                    /product="Unknown protein"
                    /protein_id="AAL75733.1"
                    /db_xref="GI:18542891"
                    /translation="MALLPCIAIAVLLSSSLAAAVTGPPGTIERVVKQQILASIPPGG
                    HGAGDVHPPVLFLTSPSGKYAAYFVRSHTVPGAGGLGADFCYVEVVVNKGGEGDAAVA
                    AGGGMSVWESECRPVSTVNTCTLLFSWHGLEVFDGSQEVWHGETNTDGTNFLQRLELV
                    DDGDMRIRDKDGELAWRASDEPRHAQHCGAPGSPGLATAFPPFAEPIGAHSSDLPFGM
                    FPGGNGRAAELPQAADAAAGALGGVGAVAPLPGAVGGAGAVAPLPGAAGGDAAAAPLP
                    GDMGGDAALAPLPGALGGVDPAFSPLPGDLPDPAAAAAAGGAAAGVGGVGAGALGAGA
                    VGAFGSQPLVDNSPYDSGAWKVDGHLVAIVVALGVVLGAI"
repeat_region       33965. .33982
                    /note="Simple Sequence Repeat (CTG)n"
gene                complement(34686. .35039)
                    /gene="OSJNBa0004E08.7"
                    /note="Hypothetical protein"
CDS                 complement(34686. .35039)
                    /gene="OSJNBa0004E08.7"
```

```
  Query Match             10.1%;  Score 40.8;  DB 8;  Length 158294;
  Best Local Similarity   71.1%;  Pred. No. 0.67;
  Matches   54;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy    166 CAGAGACCGAGGAAAACAACCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTT 225
          || |||| ||||| ||  ||||||| | ||| |||| |  |     | || ||||||||
Db  77928 CATAGACAGAGGAGAATGACCAGATGCAAGGCCCAACGATTGTTTGTTCCGTATTTCCTC 77869

Qy    226 TGACTTCAGGTGGACA 241
          ||||||||  || |||
Db  77868 TGACTTCAATTGTACA 77853


RESULT 14
E22116
LOCUS       E22116                   936 bp    DNA     linear   PAT 18-JUN-2001
DEFINITION  Yeast capable of low-temperature regulation of activation and
            methods for preparation and utilization thereof.
ACCESSION   E22116
VERSION     E22116.1  GI:13023841
KEYWORDS    JP 1999042090-A/1.
SOURCE      Saccharomyces cerevisiae.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 936)
  AUTHORS   Shun,H., Yoshinobu,K., Akio,O., Yukio,M., Tetsuji,T., Yoshie,Y.,
            Kozo,O. and Yuto,T.
  TITLE     Yeast capable of low-temperature regulation of activation and
            methods for preparation and utilization thereof
  JOURNAL   Patent: JP 1999042090-A 1 16-FEB-1999;
            SHOWA SANGYO CO LTD,KANEKA CORP
COMMENT     OS   Saccharomyces cerevisiae
            PN   JP 1999042090-A/1
            PD   16-FEB-1999
            PF   29-JUL-1997 JP 1997203652
            PR
            PI   SHUN HARASHIMA,YOSHINOBU KANEKO,AKIO OGAWA,YUKIO MUKAI, PI
            TETSUJI TOMITA,
            PI   YOSHIE YAMAMOTO,KOZO OTAKU,YUTO TAKADA
            PC   C12N15/09,A21D8/04,A21D10/02,C12N1/19//C12N9/16,(C12N1/19, PC
             C12R1:865),
            PC   C12N15/00
            CC   Strandedness: Double;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers
            FT   CDS             1. .936.
FEATURES             Location/Qualifiers
     source          1. .936
                     /organism="Saccharomyces cerevisiae"
                     /db_xref="taxon:4932"
BASE COUNT      261 a    189 c    233 g    253 t
ORIGIN

  Query Match             10.0%;  Score 40.4;  DB 6;  Length 936;
  Best Local Similarity   59.9%;  Pred. No. 0.64;
  Matches   85;  Conservative    0;  Mismatches   56;  Indels    1;  Gaps    1;

Qy      4 GGCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCT 63
          | |||||||||||| ||||    |    || | ||  ||   |  |  || || || | |
Db    703 GCCCGGGCTCACCAACTGGTTATGGAGGGGTTCAAGTACCATTTTCCTGA-AAAGGACGT 761

Qy     64 TGAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATT 123
          || |   |||||||| || ||  | |  |     |||||||| |||| ||   | || |
```

```
Db    762 TGTAACCGTGTGGTCCGCGCCAAATTACTGTTATAGATGTGGTAATGTCGCGAGTGTAAT 821

Qy    124 AAGCTTCGATGAGAAGAAGGAA 145
            |   |||||||| |   ||||
Db    822 GAAGGTCGATGAGGATCTGGAA 843


RESULT 15
YSCSIT4C
LOCUS       YSCSIT4C                1978 bp    DNA     linear   PLN 06-DEC-1994
DEFINITION  Saccharomyces cerevisiae protein phosphatase catalytic subunit
            homologue (SIT4) gene, complete cds.
ACCESSION   M24395
VERSION     M24395.1  GI:341163
KEYWORDS    SIT4 gene; homologue; protein phosphatase catalytic subunit.
SOURCE      Saccharomyces cerevisiae DNA.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 1978)
  AUTHORS   Arndt,K.T., Styles,C.A. and Fink,G.R.
  TITLE     A suppressor of a HIS4 transcriptional defect encodes a protein
            with homology to the catalytic subunit of protein phosphatases
  JOURNAL   Cell 56 (4), 527-537 (1989)
  MEDLINE   89136000
   PUBMED   2537149
FEATURES             Location/Qualifiers
     source          1. .1978
                     /organism="Saccharomyces cerevisiae"
                     /db_xref="taxon:4932"
     gene            458. .1393
                     /gene="SIT4"
     CDS             458. .1393
                     /gene="SIT4"
                     /note="homologue of protein phosphatase catalytic subunit"
                     /codon_start=1
                     /protein_id="AAA56864.1"
                     /db_xref="GI:598433"
                     /translation="MVSRGPDEWLETIKKCQALTENEMKQLCEMVKELLMEESNIQPV
                     QTPVTVCGDIHGQFHDLLELFRTAGGFPDDINYIFLGDYVDRGYYSLETFTLLMCLKV
                     KYPAKITLVRGNHESRQITQVYGFYEECLNKYGSTTVWKYCCQVFDFLTLAAIIDGKI
                     LCVHGGLSPEIRMLDQIRVLSRAQEVPHEGGFSDLLWSDPDNVEAWQVSPRGAGWLFG
                     SKVAREFNHVNGLNLIARAHQLVMEGFKYHFPEKDVVTVWSAPNYCYRCGNVASVMKV
                     DEDLEPTFKIFSAVPDDYIRESTANHNNQRAGYFL"
BASE COUNT      627 a    411 c    392 g    548 t
ORIGIN

  Query Match             10.0%;  Score 40.4;  DB 8;  Length 1978;
  Best Local Similarity   59.9%;  Pred. No. 0.67;
  Matches   85;  Conservative    0;  Mismatches   56;  Indels    1;  Gaps    1;

Qy      4 GGCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCT 63
          | |||||||||||| ||||    |    || | ||  ||   |  |  || || || | |
Db   1160 GCCCGGGCTCACCAACTGGTTATGGAGGGGTTCAAGTACCATTTTCCTGA-AAAGGACGT 1218

Qy     64 TGAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATT 123
          || |   |||||||| || ||  | |  |     |||||||| |||| ||   | || |
Db   1219 TGTAACCGTGTGGTCCGCGCCAAATTACTGTTATAGATGTGGTAATGTCGCGAGTGTAAT 1278

Qy    124 AAGCTTCGATGAGAAGAAGGAA 145
            |   |||||||| |   ||||
Db   1279 GAAGGTCGATGAGGATCTGGAA 1300


Search completed: January  2, 2003, 13:24:16
Job time : 3404 secs
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: January 2, 2003, 11:48:16 ; Search time 148 Seconds
(without alignments)
6116.914 Million cell updates/sec

Title: US-09-850-147-1
Perfect score: 402
Sequence: 1 ccgggccgggctcaccagct.....ggcataacttcctgtgttcc 402

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	223.4	55.6	1444	21	AAC46559	Zea mays DNA fragm
	2	93.6	23.3	1388	21	AAC47854	Arabidopsis thalia
	3	91.4	22.7	1310	21	AAC47706	Arabidopsis thalia
	4	91.4	22.7	1311	21	AAC38705	Arabidopsis thalia
	5	89.2	22.2	1083	21	AAC50898	Arabidopsis thalia
	6	86.8	21.6	912	24	ABN85631	Protein phosphatas
	7	40.4	10.0	936	20	AAX15651	Protein phosphatas
	8	37.2	9.3	252	14	AAQ36833	Leu8 hPTH mutein g
c	9	37.2	9.3	2413	23	ABV25283	Human prostate exp
c	10	36.4	9.1	1971	24	ABL50444	Human nGPCR-Seq102
c	11	36.4	9.1	2529	24	AAD29669	Human G-protein co
c	12	36.4	9.1	2553	24	AAD22616	Human G protein-co
c	13	36.2	9.0	349980	22	AAH41224	Pyrococcus abyssi
	14	35.6	8.9	234	14	AAQ35228	Gene for hPTH mute
	15	35.6	8.9	234	14	AAQ36836	Leu8 hPTH (7-84) m
	16	35.6	8.9	237	14	AAQ35229	Gene for hPTH mute
	17	35.6	8.9	240	14	AAQ35230	Gene for hPTH mute
	18	35.6	8.9	243	14	AAQ36829	Gene for hPTH mute
	19	35.6	8.9	252	13	AAQ24478	Parathyroid hormon
	20	35.6	8.9	252	14	AAQ36831	Gene for Cys35 hPT
	21	35.6	8.9	252	14	AAQ36832	Met18 hPTH mutein
	22	35.6	8.9	252	14	AAQ36837	Leu8, 18 hPTH mute
	23	35.6	8.9	258	24	ABL49692	Human parathyroid
	24	35.6	8.9	263	13	AAQ24479	Parathyroid hormon
	25	35.6	8.9	745	22	AAI95049	Human neuroblastom
c	26	34.8	8.7	1984	23	AAD23424	Human lung tumour-
	27	34.8	8.7	9345	23	ABL09624	Drosophila melanog
	28	34.2	8.5	543	22	AAH71196	Human cervical can
	29	34	8.5	18183	22	AAS46280	Tumour suppressor
	30	34	8.5	18183	24*	ABL70112	Chemically treated
	31	34	8.5	18183	24	ABK31159	Signal transductio
c	32	33.8	8.4	183	24	ABN91518	Staphylococcus epi
	33	33.8	8.4	438	14	AAQ60712	Human brain Expres
c	34	33.8	8.4	12677	22	AAK72352	Human immune/haema
c	35	33.8	8.4	12680	22	AAK72351	Human immune/haema
c	36	33.4	8.3	989	20	AAX13555	Enterococcus faeca
	37	33.2	8.3	1530	16	AAQ86694	Candida tropicalis
	38	33.2	8.3	3508	16	AAQ86693	Candida tropicalis
	39	32.8	8.2	1409	22	AAI60523	Human polynucleoti
c	40	32.4	8.1	51259	18	AAX83007	Partial mouse WRN
	41	32.2	8.0	642	23	ABK42562	Genomic sequence #
	42	32.2	8.0	642	23	ABK42563	Genomic sequence #
	43	32.2	8.0	642	23	ABK42564	Genomic sequence #
	44	32.2	8.0	642	23	ABK42565	Genomic sequence #
c	45	32.2	8.0	855	23	AAS51918	Staphylococcus aur

ALIGNMENTS

RESULT 1
AAC46559
ID AAC46559 standard; DNA; 1444 BP.
XX
AC AAC46559;
XX
DT 18-OCT-2000 (first entry)
XX
DE Zea mays DNA fragment SEQ ID NO: 50583.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic;
KW pathway; promoter; termination sequence; corn; ss.
XX
OS Zea mays subsp. mays.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.

PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 55.6%; Score 223.4; DB 21; Length 1444;
Best Local Similarity 79.5%; Pred. No. 1.6e-63;
Matches 314; Conservative 0; Mismatches 76; Indels 5; Gaps 4;

```
Qy       5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
          ||||||||||||||||||| |||| | |  | ||  |||||| ||||||||| |||||||
Db    1005 GCCGGGCTCACCAGCTGGTCCAGGAAGGCTTAAAGTACATGTTCCAGGAGAA-GGGCCTT 1063

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
          |    ||||||||||||||  | |  |    |||||||||||||  |||||||| |||
Db    1064 GTCACTGTGTGGTCTGCACCTAATTATTGCTACAGATGTGGCAATGTCGCTTCTATCTTA 1123

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
          |||||||||||||||| ||||||| |||||||| |||||||||||| ||||| |||||
Db    1124 AGCTTCGATGAGAAGATGGAAAGG-GATGTCAAGTTCTTCACAGAGACAGAGGAGAACAA 1182

Qy     185 CCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGACAATT 244
          |||||| | ||| |||||| |  | | ||||||||||||| ||||||||  || ||||||
Db    1183 CCAGATGCGAGGCCCAAGGAC-CGCAGTCCCATATTTCCTATGACTTCAATTGTACAATT 1241

Qy     245 TTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGG 304
          |||||||  |  ||||  |||||| ||||| |||||| ||||||||||   ||| |  |
Db    1242 TTAGAAGTC--TATTCGAATCTGTAAGGGCTAAAGATCAATCAGATGTGTTTATTGTTTT 1299

Qy     305 AGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACAAGTA 364
          |||| ||||| ||| ||| ||||||||| ||  ||||||| |||| ||| | |||| |||
Db    1300 AGGTCCAGATTATAGTGCAATCTGATGCGTTTGATTTCTGTTTTCCTTTAACCACATGTA 1359

Qy     365 CAAAAAACCAACGGGAGGGGCATAACTTCCTGTGT 399
          ||| | |||| |||||  | ||||  ||||||| |
Db    1360 CAATAGACCATCGGGACAGTCATATGTTCCTGTAT 1394
```

RESULT 2
AAC47854
ID AAC47854 standard; DNA; 1388 BP.
XX
AC AAC47854;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 55364.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.

XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 23.3%; Score 93.6; DB 21; Length 1388;
Best Local Similarity 72.3%; Pred. No. 2e-20;
Matches 162; Conservative 0; Mismatches 59; Indels 3; Gaps 3;

```
Qy       5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
           ||||  | ||||| || |  || . | | |||||  |||||| ||| || ||  ||||||
Db     823 GCCGCTCGCACCAACTTGTACAAGAAGGTCTTAAGTACATGTTCCAAGATAA-AGGCCTT 881

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
           | |  ||| ||||||||||  |    |    | | ||||| ||||  ||||||||||
Db     882 GTAACTGTATGGTCTGCACCTAATAACTGTTACCGCTGTGGGAATGTCGCTTCTATATTG 941

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
           || ||| |||| || | ||||||| || || ||| ||||||||||||| || || |||||
Db     942 AGTTTCAATGACAACATGGAAAGG-GAAGTGAAGTTCTTCACAGAGACAGAAGAGAACAA 1000

Qy     185 CCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGA 228
            || ||   |||||||||| | |||| | || ||||||| |||
Db    1001 TCAAATGAGAGGGCCAAGGAC-TGGAGTTCCGTATTTCCTATGA 1043
```

RESULT 3
AAC47706
ID AAC47706 standard; DNA; 1310 BP.
XX
AC AAC47706;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 54816.
XX

KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.

PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 22.7%; Score 91.4; DB 21; Length 1310;
Best Local Similarity 70.4%; Pred. No. 1e-19;
Matches 164; Conservative 0; Mismatches 66; Indels 3; Gaps 3;

```
Qy       5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
          | || ||||||||||| |  ||   | | || ||  |||||| ||| || ||  || |||
Db     868 GTCGCGCTCACCAGCTTGTACAAGAAGGTCTCAAGTACATGTTCCAAGATAA-AGGTCTT 926

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
           | |  |||||||||||| ||  | |  |    | | ||||| ||||  ||||||||||
Db     927 GTAACTGTGTGGTCTGCGCCTAATTACTGCTACCGTTGTGGGAATGTGGCTTCTATATTG 986

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
           || ||| |||| || | ||||||| || || ||| ||||||| ||||| || || |||||
Db     987 AGTTTCAATGACAACATGGAAAGG-GAAGTGAAGTTCTTCACGGAGACAGAAGAGAACAA 1045

Qy     185 CCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTG 237
            || ||   |||||||||| |  ||  | || |||||||| ||| |   | ||
Db    1046 TCAAATGAGAGGGCCAAGGAC-AGGTGTTCCTTATTTCCTGTGATTCATGATG 1097
```

RESULT 4
AAC38705
ID AAC38705 standard; DNA; 1311 BP.
XX
AC AAC38705;

XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 21942.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 22.7%; Score 91.4; DB 21; Length 1311;
Best Local Similarity 70.4%; Pred. No. 1e-19;
Matches 164; Conservative 0; Mismatches 66; Indels 3; Gaps 3;

```
Qy      5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
          | || |||||||||||| |  ||   | | || ||  |||||| ||| || ||  || |||
Db    868 GTCGCGCTCACCAGCTTGTACAAGAAGGTCTCAAGTACATGTTCCAAGATAA-AGGTCTT 926

Qy     65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
          | |  |||||||||||| ||  | |  |    | | ||||| ||||  ||||||||||
Db    927 GTAACTGTGTGGTCTGCGCCTAATTACTGCTACCGTTGTGGGAATGTGGCTTCTATATTG 986

Qy    125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
          || ||| |||| || | ||||||| || || ||| ||||||| ||||| || || |||||
Db    987 AGTTTCAATGACAACATGGAAAGG-GAAGTGAAGTTCTTCACGGAGACAGAAGAGAACAA 1045

Qy    185 CCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTG 237
           || ||   |||||||||| |  ||  | || |||||||| ||| |   | ||
Db   1046 TCAAATGAGAGGGCCAAGGAC-AGGTGTTCCTTATTTCCTGTGATTCATGATG 1097
```

```
RESULT 5
AAC50898
ID   AAC50898 standard; DNA; 1083 BP.
XX
AC   AAC50898;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 66541.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
```

```
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match           22.2%;  Score 89.2;  DB 21;  Length 1083;
  Best Local Similarity   70.9%;  Pred. No. 5.1e-19;
  Matches  146;  Conservative    0;  Mismatches   58;  Indels    2;  Gaps    2;

Qy       5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
           | || |||||||||||| |  ||   | | || ||  |||||| ||| || ||  || |||
Db     692 GTCGCGCTCACCAGCTTGTACAAGAAGGTCTCAAGTACATGTTCCAAGATAA-AGGTCTT 750

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
           | |  |||||||||||| ||  | |  |    | | ||||| |||  ||||||||||
Db     751 GTAACTGTGTGGTCTGCGCCTAATTACTGCTACCGTTGTGGGAATGTGGCTTCTATATTG 810

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
           || ||| |||| || | ||||||| || || ||| ||||||| ||||| || || |||||
Db     811 AGTTTCAATGACAACATGGAAAGG-GAAGTGAAGTTCTTCACGGAGACAGAAGAGAACAA 869

Qy     185 CCAGATTCCAGGGCCAAGGGCTTGGA 210
           || ||   |||||||||| |  | |
Db     870 TCAAATGAGAGGGCCAAGGACAGGTA 895


RESULT 6
ABN85631
ID   ABN85631 standard; DNA; 912 BP.
XX
AC   ABN85631;
XX
DT   13-SEP-2002  (first entry)
XX
DE   Protein phosphatase 2A catalytic subunit encoding gene.
XX
KW   Protein phosphatase 2A; enzyme; plant; flowering; crop; gene; ds.
XX
OS   Unidentified.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..912
FT                   /*tag= a
FT                   /product= "Protein phosphatase 2A catalytic subunit"
XX
PN   KR2001106331-A.
XX
PD   29-NOV-2001.
XX
PF   19-OCT-2001; 2001KR-0064613.
XX
PR   19-OCT-2001; 2001KR-0064613.
XX
PA   (KOKU-) KOREA KUMHO PETROCHEMICAL CO LTD.
XX
PI   Kang JG,  Park CM,  Song PS;
XX
DR   WPI; 2002-349593/38.
DR   P-PSDB; ABB83820.
XX
PT   Nucleic acid molecule encoding catalytic subunit of protein phosphatase
PT   2A that regulates flowering time in plants -
XX
PS   Disclosure; Fig 2; 19pp; Korean.
XX
CC   The invention relates to a nucleic acid molecule (ABN85631) encoding the
CC   catalytic subunit of protein phosphatase 2A (ABB83820) that regulates
CC   flowering timing in plants is provided, thereby delaying flowering time
CC   in plants and transferring nutrients effectively to roots, stems and
CC   leaves of crops.
XX
SQ   Sequence 912 BP; 257 A; 170 C; 207 G; 278 T; 0 other;

  Query Match           21.6%;  Score 86.8;  DB 24;  Length 912;
  Best Local Similarity   69.5%;  Pred. No. 2.9e-18;
  Matches  146;  Conservative    0;  Mismatches   62;  Indels    2;  Gaps    2;

Qy       5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
           | || || || || || |  ||   | | |||||| |||||| ||| || ||  ||||||
Db     692 GTCGAGCACATCAACTTGTACAAGAAGGCCTTAAATACATGTTCCAAGATAA-AGGCCTT 750

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
           | |  ||| ||||||||||  | |  |    | | ||||| |||| |||||||||  |
Db     751 GTAACTGTATGGTCTGCACCTAATTACTGTTACCGTTGTGGAAATGTAGCTTCTATTCTT 810

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
           || ||| |||| || | ||| | |||| || ||| | ||||| || || ||||| |||||
Db     811 AGTTTCAATGAAAATATGGAGA-GAGAAGTTAAGTTTTTCACTGAAACAGAGGAGAACAA 869

Qy     185 CCAGATTCCAGGGCCAAGGGCTTGGAATCC 214
           |||||   |||||||||| |  |   |||
Db     870 TCAGATGAGAGGGCCAAGGACAGGTGTTCC 899
```

```
RESULT 7
AAX15651
ID   AAX15651 standard; DNA; 936 BP.
XX
AC   AAX15651;
XX
DT   07-MAY-1999  (first entry)
XX
DE   Protein phosphatase gene coding sequence of Saccharomyces cerevisiae.
XX
KW   Protein phosphatase gene; growth; fermentation activity;
KW   dough production; yeast; ds.
XX
OS   Saccharomyces cerevisiae.
XX
PN   JP11042090-A.
XX
PD   16-FEB-1999.
XX
PF   29-JUL-1997;   97JP-0203652.
XX
PR   29-JUL-1997;   97JP-0203652.
XX
PA   (KANF ) KANEKA CORP.
PA   (SHOS ) SHOWA SANGYO CO.
XX
DR   WPI; 1999-197822/17.
XX
PT   New yeast of controlled activation at low temperatures - useful for
PT   improving the quality of dough
XX
PS   Claim 3; Page 16-17; 41pp; Japanese.
XX
CC   The present sequence represents a protein phosphatase gene sequence
CC   of Saccharomyces cerevisiae. The specification describes new
CC   S. cerevisiae in which the growth and/or the fermentation activity
CC   is controlled at least in the range of 0-20 degrees Celsius. These
CC   yeast are prepared by deleting the function of at least one protein
CC   phosphatase gene. The yeast is useful in the production of dough.
XX
SQ   Sequence 936 BP; 261 A; 189 C; 233 G; 253 T; 0 other;

  Query Match             10.0%;  Score 40.4;  DB 20;  Length 936;
  Best Local Similarity   59.9%;  Pred. No. 0.0077;
  Matches   85;  Conservative    0;  Mismatches   56;  Indels    1;  Gaps    1;

Qy       4 GGCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCT 63
           | ||||||||||||| ||||    |    || | ||  ||   |  |  || || || | |
Db     703 GCCCGGGCTCACCAACTGGTTATGGAGGGGTTCAAGTACCATTTTCCTGA-AAAGGACGT 761

Qy      64 TGAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATT 123
           || |   |||||||| || ||  | |  |     |||||||| ||||  ||   | || |
Db     762 TGTAACCGTGTGGTCCGCGCCAAATTACTGTTATAGATGTGGTAATGTCGCGAGTGTAAT 821

Qy     124 AAGCTTCGATGAGAAGAAGGAA 145
             |   |||||||| |   ||||
Db     822 GAAGGTCGATGAGGATCTGGAA 843


RESULT 8
AAQ36833
ID   AAQ36833 standard; DNA; 252 BP.
XX
AC   AAQ36833;
XX
DT   09-JUN-1993  (first entry)
XX
DE   Leu8 hPTH mutein gene.
XX
KW   Human parathyroid hormone; hPTH; antagonist; hypoparathyroidism;
KW   hypercalcaemia; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..252
FT                   /*tag=  a
FT   mutation        19..24
FT                   /*tag=  b
FT                   /note= "TTA ATG Leu-Met -> CTG CTG Leu-Leu"
XX
PN   EP528271-A.
XX
PD   24-FEB-1993.
XX
PF   05-AUG-1992;   92EP-0113322.
XX
PR   07-AUG-1991;   91JP-0198056.
PR   26-JUN-1992;   92JP-0169713.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Fukuda T;
XX
DR   WPI; 1993-060187/08.
DR   P-PSDB; AAR30858.
XX
PT   New human parathyroid mutein(s) - useful for treating e.g.
PT   osteoporosis, hypoparathyroidism, hyperparathyroidism,
PT   hypercalcaemia, hypertension etc.
XX
PS   Example; Page 21; 88pp; English.
XX
CC   The sequence is that of a gene coding for human parathyroid
CC   hormone (hPTH) where Met8 is substd. by Leu. The peptide
CC   can have higher stability, enhanced activity and improved
CC   absorption by tissues. It can act as a PTH antagonist and
CC   can be used as a therapeutic agent for hypercalcaemia and
CC   hyperparathyroidism.
XX
SQ   Sequence 252 BP; 72 A; 53 C; 65 G; 62 T; 0 other;

  Query Match             9.3%;  Score 37.2;  DB 14;  Length 252;
  Best Local Similarity   56.6%;  Pred. No. 0.046;
  Matches   69;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
           ||| |||   | | |||||  |  |  |  || || ||  ||| |||  | ||   ||
Db      78 GAAGTTGCAGGATGTGCACAATTTTGTTGCCTTAGGTGCCGCATTGGCTCCTCGTGATGC 137

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
             | |||    |||  | | |||   || | ||| |||||    |||| | | |||||
Db     138 TGGTTCCCAAAGACCACGTAAAAAGGAAGACAATGTCTTAGTTGAGAGCCATGAAAAATC 197

Qy     185 CC 186
           ||
Db     198 CC 199


RESULT 9
ABV25283/c
ID   ABV25283 standard; cDNA; 2413 BP.
XX
AC   ABV25283;
XX
DT   16-SEP-2002  (first entry)
XX
DE   Human prostate expression marker cDNA 25274.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200160860-A2.
XX
```

PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 4968; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ Sequence 2413 BP; 565 A; 651 C; 612 G; 576 T; 9 other;

Query Match 9.3%; Score 37.2; DB 23; Length 2413;
Best Local Similarity 50.6%; Pred. No. 0.14;
Matches 90; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

Qy 217 TATTTCCTTTGACTTCAGGTGGACAATTTTAGAAGGCAATATTTAAATCTGGAAGGGGTA 276
Db 2235 TATTTTTTTTTTTTACCAGTAGTTCCTTTCCCGCTTTATTTTTTAGCTGCTTTTTGGGTT 2176

Qy 277 AAGAACAATCAGATGGTGTTAATTGTGGAGGTTCAGATCATAATGCCATCTGATGCATTC 336
Db 2175 TTATACAATGAACATGTATTAATTGTAGAAGAAAACGATGTCATCCTTTATGATAAAATC 2116

Qy 337 CATTTCTGGTTTCTTTTGAACACAAGTACAAAAAACCAACGGGAGGGGCATAACTTCC 394
Db 2115 CATTTCCATTTTAGCTTTTTTAAAAAAACAAAAAGCTGTTGTGGACAGATGAACATCC 2058

RESULT 10
ABL50444/c
ID ABL50444 standard; cDNA; 1971 BP.
XX
AC ABL50444;
XX
DT 18-JUN-2002 (first entry)
XX
DE Human nGPCR-Seq1025 polynucleotide sequence SEQ ID NO:15.
XX
KW Human; G protein-coupled receptor; GPCR; nGPCR; vulnerary; antiasthmatic;
KW anti-HIV; analgesic; cytostatic; antidiabetic; anorectic; metabolic;
KW hypertensive; hypotensive; thrombolytic; antiparkinsonian; cardiant;
KW antiatherosclerotic; neuroleptic; antimigraine; neuroprotective; cancer;
KW tranquilliser; antidepressant; antiinflammatory; antirheumatic; diabetes;
KW antiarthritic; immunosuppressive; antipsoriatic; viral infection; pain;
KW human immunodeficiency virus; obesity; anorexia; hypotension; thrombosis;
KW hypertension; myocardial infarction; cardiomyopathy; atherosclerosis;
KW Parkinson's disease; schizophrenia; migraine; anxiety; manic depression;
KW dementia; Huntington's disease; thyroid disorder; inflammatory condition;
KW rheumatoid arthritis; autoimmune disorder; hormonal disorder; psoriasis;
KW renal failure; movement disorder; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200183553-A2.
XX
PD 08-NOV-2001.
XX
PF 01-MAY-2001; 2001WO-US14050.
XX
PR 03-MAY-2000; 2000US-201548P.
PR 03-MAY-2000; 2000US-201549P.
PR 03-MAY-2000; 2000US-201550P.
PR 03-MAY-2000; 2000US-201551P.
PR 03-MAY-2000; 2000US-201552P.
PR 03-MAY-2000; 2000US-201632P.
PR 03-MAY-2000; 2000US-201633P.
PR 03-MAY-2000; 2000US-201978P.
XX
PA (PHAA) PHARMACIA & UPJOHN CO.
XX
PI Parodi LA, Lind P, Sejlitz T;
XX
DR WPI; 2002-041488/05.
DR P-PSDB; ABB06813.
XX
PT Novel G protein-coupled receptor-x nucleic acid molecule and
PT polypeptides encoded by them, useful for treating cancers, psoriasis,
PT Alzheimer's disease, atherosclerosis, rheumatoid arthritis, obesity,
PT anorexia -
XX
PS Claim 3; Page 70-71; 193pp; English.
XX
CC ABL50430 to ABL50496 encode the human novel G protein-coupled receptor
CC (nGPCR) proteins given in ABB06799 to ABB06867. The nGPCR sequences have
CC vulnerary, antiasthmatic, anti-HIV, analgesic, cytostatic, antidiabetic,
CC anorectic, metabolic, hypertensive, hypotensive, antiatherosclerotic,
CC thrombolytic, antiparkinsonian, cardiant, neuroleptic, antimigraine,
CC neuroprotective, tranquilliser, antidepressant, antiinflammatory,
CC antirheumatic, antiarthritic, immunosuppressive and antipsoriatic
CC activities. They can be used as immune response inducers, nGPCR
CC expression or activity modulators and in gene therapy. nGPCR
CC polynucleotide sequences can be used as probes to screen databases and
CC nucleic acid libraries, for screening restriction fragment length
CC polymorphism associated with certain disorders, as well as for genetic
CC mapping. nGPCR sequences can be used in the treatment of diseases such
CC as viral infections caused by human immunodeficiency virus (HIV)-1 or
CC HIV-2, pain, cancers, diabetes, obesity, anorexia, hypotension,
CC hypertension, thrombosis, myocardial infarction, cardiomyopathies,
CC atherosclerosis, Parkinson's disease, schizophrenia, migraine, anxiety,
CC manic depression, dementia, Huntington's disease, thyroid disorders,
CC inflammatory conditions, rheumatoid arthritis, autoimmune disorders,
CC hormonal disorders, renal failure, psoriasis, movement disorders.
XX
SQ Sequence 1971 BP; 652 A; 311 C; 374 G; 634 T; 0 other;

Query Match 9.1%; Score 36.4; DB 24; Length 1971;
Best Local Similarity 54.5%; Pred. No. 0.24;
Matches 73; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 151 ATGTCAAGGTCTTCACAGAGACCGAGGAAAACAACCAGATTCCAGGGCCAAGGGCTTGGA 210
Db 246 ATAACAAGCTCTGGAAGAGTACAGAAGAATGACTACAGAGCCCACGACGACAGTCTGGCA 187

Qy 211 ATCCCATATTTCCTTTGACTTCAGGTGGACAATTTTAGAAGGCAATATTTAAATCTGGAA 270
Db 186 CAGAGACATACCCTTTGATTCGAGGTAGACATTAGCTGATGACATTTTTGTAATTTTGCA 127

```
Qy    271 GGGGTAAAGAACAA 284
              ||||   || |
Db    126 TATGTAACACACTA 113


RESULT 11
AAD29669/c
ID   AAD29669 standard; cDNA; 2529 BP.
XX
AC   AAD29669;
XX
DT   17-MAY-2002  (first entry)
XX
DE   Human G-protein coupled receptor (GCREC-3) cDNA.
XX
KW   Human; G-protein coupled receptor; GCREC-3; cell proliferative disorder;
KW   neurological; cardiovascular; gastrointestinal; autoimmune; inflammatory;
KW   metabolic; hepatitis; psoriasis; cancer; epilepsy; Alzheimer's disease;
KW   Pick's disease; Huntington's disease; Parkinson's disease; hypertension;
KW   atherosclerosis; myocardial infarction; gastritis; cirrhosis; cytostatic;
KW   osteoporosis; Crohn's disease; acquired immunodeficiency syndrome; AIDS;
KW   anaemia; asthma; rheumatoid arthritis; diabetes; obesity; drug screening;
KW   transgenic animal; allergy; gene therapy; hepatotropic; anticonvulsant;
KW   nootropic; neuroprotective; cardiant; immunosuppressive; anorectic;
KW   virucide; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             24..1310
FT                   /*tag=  a
FT                   /product= "Human GCREC-3"
XX
PN   WO200210387-A2.
XX
PD   07-FEB-2002.
XX
PF   25-JUL-2001; 2001WO-US23433.
XX
PR   27-JUL-2000; 2000US-221478P.
PR   03-AUG-2000; 2000US-223268P.
PR   21-AUG-2000; 2000US-227054P.
PR   08-SEP-2000; 2000US-231121P.
PR   13-SEP-2000; 2000US-232243P.
PR   15-SEP-2000; 2000US-232691P.
PR   22-SEP-2000; 2000US-235146P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Thornton M,  Patterson C,  Lal P,  Burford N,  Yue H,  Gandhi AR;
PI   Elliot VS,  Ramkumar J,  Baughn MR,  Kallick DA,  Walia NK;
PI   Hafalia AJA,  Yao MG,  Lu Y,  Tribouley CM,  Policky JL,  Kearney L;
PI   Graul RC,  Warren BA,  Lee EA,  Ding L;
XX
DR   WPI; 2002-188744/24.
DR   P-PSDB; AAE18642.
XX
PT   New human G-protein coupled receptor polypeptide for diagnosis,
PT   prevention and treatment of cell proliferative, neurological,
PT   cardiovascular, gastrointestinal, autoimmune/inflammatory, and
PT   metabolic disorders -
XX
PS   Claim 5; Page 138-139; 150pp; English.
XX
CC   The invention relates to novel human G-protein coupled receptors
CC   (GCREC) and their encoding polynucleotides. GCREC is useful as an
CC   immunogen for preparing monoclonal and polyclonal antibodies. GCREC is
CC   useful for diagnosing, treating and preventing a cell proliferative
CC   disorder (e.g., hepatitis, psoriasis, cancer), a neurological disorder
CC   (e.g., epilepsy, Alzheimer's disease, Pick's disease, Huntington's
CC   disease, Parkinson's disease), a cardiovascular disorder (e.g.,
CC   atherosclerosis, hypertension, myocardial infarction), gastrointestinal
CC   disorder (e.g., gastritis, cirrhosis, Crohn's disease), an autoimmune/
CC   inflammatory disorder (e.g., acquired immunodeficiency syndrome (AIDS),
CC   allergy, anaemia, asthma, rheumatoid arthritis), a metabolic disorder
CC   (e.g., diabetes, obesity, osteoporosis), and viral infections. GCREC is
CC   useful in a number of drug screening techniques, and to analyse the
CC   proteome of a tissue or cell type. GCREC is useful for creating knockin
CC   humanised animals or transgenic animals to model human diseases, in
CC   somatic or germline gene therapy, to generate a transcript image of a
CC   tissue or cell type, for detecting differences in the chromosomal
CC   location due to translocation, inversion, etc., among normal, carrier
CC   or affected individuals, and as hybridization probes for mapping
CC   naturally occurring genomic sequences. GCREC is useful in Southern or
CC   northern analysis, dot blot or other membrane-based technologies, in PCR
CC   technologies, in dipstick, pin, multiformat enzyme linked immunosorbant
CC   (ELISA)-like assays, and in microarrays utilising fluids or tissues from
CC   patients to detect altered GCREC expression. The present sequence is
CC   human GCREC-3 cDNA.
XX
SQ   Sequence 2529 BP; 726 A; 516 C; 525 G; 762 T; 0 other;

  Query Match             9.1%;  Score 36.4;  DB 24;  Length 2529;
  Best Local Similarity   54.5%;  Pred. No. 0.27;
  Matches   73;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy    151 ATGTCAAGGTCTTCACAGAGACCGAGGAAAACAACCAGATTCCAGGGCCAAGGGCTTGGA 210
          ||  |||| |||  |      || || |||      ||||  ||| | | |  | || | |
Db    809 ATAACAAGCTCTGGAAGAGTACAGAAGAATGACTACAGAGCCCACGACGACAGTCTGGCA 750

Qy    211 ATCCCATATTTCCTTTGACTTCAGGTGGACAATTTTAGAAGGCAATATTTAAATCTGGAA 270
              | ||  ||||||| |  |||| |||| |    || | || | ||  ||| | | |
Db    749 CAGAGACATACCCTTTGATTCGAGGTAGACATTAGCTGATGACATTTTTGTAATTTTGCA 690

Qy    271 GGGGTAAAGAACAA 284
            ||||   || |
Db    689 TATGTAACACACTA 676


RESULT 12
AAD22616/c
ID   AAD22616 standard; cDNA; 2553 BP.
XX
AC   AAD22616;
XX
DT   26-FEB-2002  (first entry)
XX
DE   Human G protein-coupled receptor (GPCR) nGPCR-1025 cDNA.
XX
KW   Human; G protein-coupled receptor; GPCR; gene therapy; cancer;
KW   tumour growth; thyroid disorder; myxoedema; renal failure;
KW   inflammatory condition; Crohn's disease; cell differentiation;
KW   homeostasis; rheumatoid arthritis; autoimmune disorder;
KW   movement disorder; central nervous system disorder; stroke;
KW   Huntington's disease; Tourette's Syndrome; Parkinson's disease;
KW   Alzheimer's disease; viral infection; human immunodeficiency virus;
KW   HIV; metabolic disorder; cardiovascular disease; type 2 diabetes;
KW   obesity; hypotension; hypertension; thrombosis; myocardial infarction;
KW   atherosclerosis; proliferative disease; hyperproliferative disorder;
KW   psoriasis; hormonal disorder; polycystic ovarian syndrome; alopecia;
KW   sexual dysfunction; respiratory ailment; asthma; brain injury;
KW   nootropic; neuroprotective; cytostatic; antirheumatic; antiarthritic;
KW   cerebroprotective; anticonvulsant; neuroleptic; anorectic; cardiant;
KW   thrombolytic; antiarteriosclerotic; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             40..1329
FT                   /*tag=  a
FT                   /product= "Human nGPCR-1025 protein"
XX
PN   WO200179292-A2.
XX
PD   25-OCT-2001.
```

```
XX
PF   17-APR-2001; 2001WO-US12470.
XX
PR   17-APR-2000; 2000US-198090P.
XX
PA   (PHAA ) PHARMACIA & UPJOHN CO.
XX
PI   Soderberg C,  Lind P;
XX
DR   WPI; 2002-026013/03.
DR   P-PSDB; AAE13533.
XX
PT   Novel isolated nucleic acid molecule encoding G protein-coupled
PT   receptor polypeptide, nGPCR-1025, useful for treating, e.g., cancer,
PT   Crohn's disease, rheumatoid arthritis, Alzheimer's disease, stroke,
PT   thrombosis and psoriasis -
XX
PS   Claim 3; Page 59-60; 98pp; English.
XX
CC   The present invention relates to an isolated nucleic acid molecule
CC   comprising a nucleotide sequence that encodes a G protein-coupled
CC   receptor (GPCR) polypeptide, nGPCR-1025. nGPCR-1025 DNA may also be
CC   useful in gene therapy. The invention is used to treat or prevent
CC   unregulated cellular growth, such as cancer cell and tumour growth,
CC   and for treating thyroid disorders (e.g., myxoedema), renal failure,
CC   inflammatory conditions (e.g., Crohn's disease), diseases related to
CC   cell differentiation and homeostasis, rheumatoid arthritis, autoimmune
CC   disorders, movement disorders, central nervous system disorders (e.g.,
CC   stroke, Huntington's disease, Tourette's Syndrome, Parkinson's disease,
CC   Alzheimer's disease), infections, such as viral infections caused by
CC   HIV-1 or HIV-2 (human immunodeficiency virus), metabolic and
CC   cardiovascular diseases and disorders (e.g., type 2 diabetes, obesity,
CC   hypotension, hypertension, thrombosis, myocardial infarction,
CC   atherosclerosis), proliferative diseases and cancers, hyperproliferative
CC   disorders (such as psoriasis), hormonal disorders (e.g., polycystic
CC   ovarian syndrome, alopecia), and sexual dysfunction. nGPCR-1025 is useful
CC   for treating respiratory ailments such as asthma and for treating acute
CC   and/or traumatic brain injury. The present sequence is a human nGPCR-1025
CC   DNA.
XX
SQ   Sequence 2553 BP; 717 A; 540 C; 532 G; 764 T; 0 other;

  Query Match             9.1%;  Score 36.4;  DB 24;  Length 2553;
  Best Local Similarity   54.5%;  Pred. No. 0.27;
  Matches   73;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy    151 ATGTCAAGGTCTTCACAGAGACCGAGGAAAACAACCAGATTCCAGGGCCAAGGGCTTGGA 210
          ||  |||| |||  |     || || |||      ||||  ||| | | |  | || | |
Db    828 ATAACAAGCTCTGGAAGAGTACAGAAGAATGACTACAGAGCCCACGACGACAGTCTGGCA 769

Qy    211 ATCCCATATTTCCTTTGACTTCAGGTGGACAATTTTAGAAGGCAATATTTAAATCTGGAA 270
            | ||  ||||||| |  |||| ||||'|    || | || | ||  ||| | | |
Db    768 CAGAGACATACCCTTTGATTCGAGGTAGACATTAGCTGATGACATTTTTGTAATTTTGCA 709

Qy    271 GGGGTAAAGAACAA 284
            ||||   || |
Db    708 TATGTAACACACTA 695


RESULT 13
AAH41224/c
ID   AAH41224 standard; DNA; 349980 BP.
XX
AC   AAH41224;
XX
DT   29-OCT-2001  (first entry)
XX
DE   Pyrococcus abyssi genomic fragment #3.
XX
KW   Hyperthermophilic archaeon; hyperthermophilic protein; ds.
XX
OS   Pyrococcus abyssi.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1..49980
FT                   /*tag=  a
FT                   /note= "This sequence overlaps with the 3' end of
FT                   AAH41223"
FT   misc_feature    300001..349980
FT                   /*tag=  b
FT                   /note= "This sequence overlaps with the 5' end of
FT                   AAH41225"
XX
PN   FR2792651-A1.
XX
PD   27-OCT-2000.
XX
PF   21-APR-1999;   99FR-0005034.
XX
PR   21-APR-1999;   99FR-0005034.
XX
PA   (CNRS ) CNRS CENT NAT RECH SCI.
PA   (IFRE-) IFREMER INST FR RECH EXPL MER.
XX
PI   Forterre P,  Thierry JC,  Prieur D,  Dietrich J,  Lecompte O;
PI   Querellou J,  Weissenbach J,  Saurin W,  Heilig R;
XX
DR   WPI; 2001-126236/14.
XX
PT   New nucleotide sequences isolated from Pyrococcus abyssi encode
PT   proteins useful in industry -
XX
PS   Claim 1; Page 347-443; 1657pp; French.
XX
CC   The present invention relates to the genomic sequence of Pyrococcus
CC   abyssi and P. abyssi proteins (see AAB96053-AAB96842). P. abyssi is a
CC   hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal
CC   vents. The present sequence is a fragment of the genomic sequence of P.
CC   abyssi. The 5' end of this sequence overlaps with the 3' end of AAH41223
CC   and the 3' end of this sequence overlaps with the 5' end of AAH41225. The
CC   proteins of the present invention have various potential industrial uses,
CC   since the proteins are stable at very high temperatures, some up to 110
CC   degrees centigrade.
CC   Note: This patent is in the same patent family as WO200065062, which
CC   contains additional sequences as shown in AAB99132-AAB99143,
CC   AAH75903-AAH75920 and AAG66436.
XX
SQ   Sequence 349980 BP; 94090 A; 78692 C; 78319 G; 98879 T; 0 other;

  Query Match             9.0%;  Score 36.2;  DB 22;  Length 349980;
  Best Local Similarity   60.8%;  Pred. No. 3.8;
  Matches   59;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy     85 CTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTAAGCTTCGATGAGAAGAAGGA 144
          | ||   |||  || ||| || ||| ||| | ||     ||| ||  | ||  |||||
Db  54308 CGAAGATTTGGCCAAATGGGGAAATAGAGTTCCTGAGGAAAGATTACAGGAAGAGAAGTG 54249

Qy    145 AAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAA 181
          | | ||| ||  |||||  |    | | |||||||||
Db  54248 ATGAAGAAGTTGAGGTCGACGATAAAATCGAGGAAAA 54212


RESULT 14
AAQ35228
ID   AAQ35228 standard; DNA; 234 BP.
XX
AC   AAQ35228;
XX
DT   09-JUN-1993  (first entry)
XX
DE   Gene for hPTH mutein lacking 6 N-terminal amino acids.
XX
KW   Human parathyroid hormone; hPTH; antagonist; hypoparathyroidism;
KW   hypercalcaemia; ss.
XX
OS   Synthetic.
```

```
XX
FH   Key             Location/Qualifiers
FT   CDS             1..234
FT                   /*tag=  a
FT                   /note= "codes for peptide lacking N-terminal region"
XX
PN   EP528271-A.
XX
PD   24-FEB-1993.
XX
PF   05-AUG-1992;   92EP-0113322.
XX
PR   07-AUG-1991;   91JP-0198056.
PR   26-JUN-1992;   92JP-0169713.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Fukuda T;
XX
DR   WPI; 1993-060187/08.
DR   P-PSDB; AAR30852.
XX
PT   New human parathyroid mutein(s) - useful for treating e.g.
PT   osteoporosis, hypoparathyroidism, hyperparathyroidism,
PT   hypercalcaemia, hypertension etc.
XX
PS   Disclosure; Page 7; 88pp; English.
XX
CC   The sequence is that of a gene coding for a human parathyroid hormone
CC   (hPTH) peptide lacking 6 hPTH N-terminal amino acids. This peptide
CC   can have higher stability, enhanced activity and improved absorption
CC   by tissues. It can act as a PTH antagonist and can be used as a
CC   therapeutic agent for hypercalcaemia and hyperparathyroidism.
XX
SQ   Sequence 234 BP; 71 A; 48 C; 57 G; 58 T; 0 other;

  Query Match             8.9%;  Score 35.6;  DB 14;  Length 234;
  Best Local Similarity   55.7%;  Pred. No. 0.15;
  Matches   68;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
           ||| |||    | | |||||  |  |  |  || || ||   || |||  | ||   ||
Db      60 GAAGTTGCAGGATGTGCACAATTTTGTTGCCTTAGGTGCCCCATTGGCTCCTCGTGATGC 119

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
           | |||    |||  | | |||   || | ||| |||||    |||| | | |||||
Db     120 TGGTTCCCAAAGACCACGTAAAAAGGAAGACAATGTCTTAGTTGAGAGCCATGAAAAATC 179

Qy     185 CC 186
           ||
Db     180 CC 181


RESULT 15
AAQ36836
ID   AAQ36836 standard; DNA; 234 BP.
XX
AC   AAQ36836;
XX
DT   09-JUN-1993  (first entry)
XX
DE   Leu8 hPTH (7-84) mutein gene.
XX
KW   Human parathyroid hormone; hPTH; antagonist; hypoparathyroidism;
KW   hypercalcaemia; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..234
FT                   /*tag=  a
FT                   /note= "encodes mutein hPTH amino acids 7-84"
FT   mutation        4..6
FT                   /*tag=  b
FT                   /note= "ATG Met -> CTC Leu"
XX
PN   EP528271-A.
XX
PD   24-FEB-1993.
XX
PF   05-AUG-1992;   92EP-0113322.
XX
PR   07-AUG-1991;   91JP-0198056.
PR   26-JUN-1992;   92JP-0169713.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Fukuda T;
XX
DR   WPI; 1993-060187/08.
DR   P-PSDB; AAR30859.
XX
PT   New human parathyroid mutein(s) - useful for treating e.g.
PT   osteoporosis, hypoparathyroidism, hyperparathyroidism,
PT   hypercalcaemia, hypertension etc.
XX
PS   Example; Page 21; 88pp; English.
XX
CC   The sequence is that of a gene coding for human parathyroid
CC   hormone (hPTH) amino acids 7-84 where Met8 is substd. by Leu.
CC   The peptide can have higher stability, enhanced activity and
CC   improved absorption by tissues. It can act as a PTH antagonist
CC   and can be used as a therapeutic agent for hypercalcaemia and
CC   hyperparathyroidism.
XX
SQ   Sequence 234 BP; 70 A; 50 C; 57 G; 57 T; 0 other;

  Query Match             8.9%;  Score 35.6;  DB 14;  Length 234;
  Best Local Similarity   55.7%;  Pred. No. 0.15;
  Matches   68;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
           ||| |||    | | |||||  |  |  |  || || ||   || |||  | ||   ||
Db      60 GAAGTTGCAGGATGTGCACAATTTTGTTGCCTTAGGTGCCCCATTGGCTCCTCGTGATGC 119

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
           | |||    |||  | | |||   || | ||| |||||    |||| | | |||||
Db     120 TGGTTCCCAAAGACCACGTAAAAAGGAAGACAATGTCTTAGTTGAGAGCCATGAAAAATC 179

Qy     185 CC 186
           ||
Db     180 CC 181


Search completed: January  2, 2003, 12:29:55
Job time : 304 secs
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: January 2, 2003, 12:23:21 ; Search time 74 Seconds
(without alignments)
1666.001 Million cell updates/sec

Title: US-09-850-147-1
Perfect score: 402
Sequence: 1 ccgggccgggctcaccagct.....ggcataacttcctgtgttcc 402

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	35.6	8.9	234	2	US-08-733-446-19	Sequence 19, Appl
	2	35.6	8.9	234	2	US-08-733-446-61	Sequence 61, Appl
	3	35.6	8.9	237	2	US-08-733-446-20	Sequence 20, Appl
	4	35.6	8.9	240	2	US-08-733-446-21	Sequence 21, Appl
	5	35.6	8.9	243	2	US-08-733-446-22	Sequence 22, Appl
	6	35.6	8.9	245	2	US-08-733-446-24	Sequence 24, Appl
c	7	35.6	8.9	247	2	US-08-733-446-25	Sequence 25, Appl
	8	35.6	8.9	248	2	US-08-733-446-26	Sequence 26, Appl
c	9	35.6	8.9	250	2	US-08-733-446-27	Sequence 27, Appl
	10	35.6	8.9	251	2	US-08-733-446-28	Sequence 28, Appl
	11	35.6	8.9	252	1	US-08-689-190-1	Sequence 1, Appli
	12	35.6	8.9	252	2	US-08-733-446-23	Sequence 23, Appl
	13	35.6	8.9	252	2	US-08-733-446-56	Sequence 56, Appl
	14	35.6	8.9	252	2	US-08-733-446-57	Sequence 57, Appl
	15	35.6	8.9	252	2	US-08-733-446-58	Sequence 58, Appl
	16	35.6	8.9	252	2	US-08-733-446-62	Sequence 62, Appl
	17	35.6	8.9	252	2	US-08-835-231-34	Sequence 34, Appl
	18	35.6	8.9	252	4	US-09-108-661-34	Sequence 34, Appl
c	19	35.6	8.9	253	2	US-08-733-446-29	Sequence 29, Appl
	20	35.6	8.9	254	2	US-08-733-446-30	Sequence 30, Appl
c	21	35.6	8.9	256	2	US-08-733-446-31	Sequence 31, Appl
	22	35.6	8.9	263	1	US-08-689-190-3	Sequence 3, Appli
	23	35.6	8.9	263	2	US-08-733-446-32	Sequence 32, Appl
c	24	35.6	8.9	265	1	US-08-689-190-4	Sequence 4, Appli
c	25	35.6	8.9	265	2	US-08-733-446-33	Sequence 33, Appl
c	26	35.2	8.8	7218	1	US-08-232-463-14	Sequence 14, Appl
c	27	33.8	8.4	183	4	US-09-134-001C-981	Sequence 981, App
c	28	32.4	8.1	51259	3	US-08-781-891-209	Sequence 209, App
	29	32	8.0	812	1	US-07-612-674-3	Sequence 3, Appli
	30	32	8.0	1696	1	US-07-612-674-1	Sequence 1, Appli
c	31	31.6	7.9	5855	1	US-08-592-214A-20	Sequence 20, Appl
c	32	31.6	7.9	5855	3	US-09-149-976-20	Sequence 20, Appl
	33	31	7.7	29793	4	US-09-302-812-38	Sequence 38, Appl
	34	31	7.7	29793	4	US-09-511-477-38	Sequence 38, Appl
	35	31	7.7	29793	4	US-09-511-507-38	Sequence 38, Appl
c	36	30.6	7.6	432	4	US-08-905-223-134	Sequence 134, App
c	37	30.4	7.6	2606	2	US-08-568-459A-7	Sequence 7, Appli
c	38	30.4	7.6	2606	2	US-08-487-826B-7	Sequence 7, Appli
c	39	30.4	7.6	2606	4	US-09-210-288-7	Sequence 7, Appli
c	40	30	7.5	1669	4	US-09-026-039-1	Sequence 1, Appli
c	41	29.6	7.4	931	4	US-09-453-702B-112	Sequence 112, App
	42	29.6	7.4	1609	4	US-09-312-285-1	Sequence 1, Appli
	43	29.6	7.4	1609	4	US-09-728-764-1	Sequence 1, Appli
	44	29.6	7.4	14672	4	US-08-961-527-111	Sequence 111, App
	45	29.4	7.3	1557	4	US-09-134-001C-1614	Sequence 1614, Ap

ALIGNMENTS

RESULT 1
US-08-733-446-19
; Sequence 19, Application US/08733446
; Patent No. 5856138
; GENERAL INFORMATION:
; APPLICANT: FUKUDA, Tsunehiko
; TITLE OF INVENTION: HUMAN PARATHYROID HORMONE MUTEINS AND
; TITLE OF INVENTION: PRODUCTION THEREOF
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
; ADDRESSEE: CUSHMAN
; STREET: 130 Water Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/733,446
; FILING DATE: 18-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/402,970
; FILING DATE:
; APPLICATION NUMBER: US/07/926,787
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: NEUNER, George W
; REGISTRATION NUMBER: 26964
; REFERENCE/DOCKET NUMBER: 42025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)523-3400
; TELEFAX: (617)523-6440
; TELEX: 200291 STRE UR
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid, synthetic DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..234
; IDENTIFICATION METHOD: E

```
US-08-733-446-19

  Query Match             8.9%;  Score 35.6;  DB 2;  Length 234;
  Best Local Similarity   55.7%;  Pred. No. 0.036;
  Matches   68;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
           ||| |||   | | |||||  |  |  |  || || ||   || ||| | ||   ||
Db      60 GAAGTTGCAGGATGTGCACAATTTTGTTGCCTTAGGTGCCCCATTGGCTCCTCGTGATGC 119

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
            | |||    |||  | | |||   || | ||| |||||    |||| | | |||||
Db     120 TGGTTCCCAAAGACCACGTAAAAAGGAAGACAATGTCTTAGTTGAGAGCCATGAAAAATC 179

Qy     185 CC 186
           ||
Db     180 CC 181


RESULT 2
US-08-733-446-61
; Sequence 61, Application US/08733446
; Patent No. 5856138
;  GENERAL INFORMATION:
;    APPLICANT:  FUKUDA, Tsunehiko
;    TITLE OF INVENTION:  HUMAN PARATHYROID HORMONE MUTEINS AND
;    TITLE OF INVENTION:  PRODUCTION THEREOF
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
;      ADDRESSEE:  CUSHMAN
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  US
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/733,446
;      FILING DATE:  18-OCT-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/402,970
;      FILING DATE:
;      APPLICATION NUMBER:  US/07/926,787
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  NEUNER, George W
;      REGISTRATION NUMBER:  26964
;      REFERENCE/DOCKET NUMBER:  42025
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)523-3400
;      TELEFAX:  (617)523-6440
;      TELEX:  200291 STRE UR
;  INFORMATION FOR SEQ ID NO:  61:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  234 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid, synthetic DNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..234
;      NAME/KEY:  mutation
;      LOCATION:  4,6
;      IDENTIFICATION METHOD:  E
US-08-733-446-61

  Query Match             8.9%;  Score 35.6;  DB 2;  Length 234;
  Best Local Similarity   55.7%;  Pred. No. 0.036;
  Matches   68;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
           ||| |||   | | |||||  |  |  |  || || ||   || ||| | ||   ||
Db      60 GAAGTTGCAGGATGTGCACAATTTTGTTGCCTTAGGTGCCCCATTGGCTCCTCGTGATGC 119

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
            | |||    |||  | | |||   || | ||| |||||    |||| | | |||||
Db     120 TGGTTCCCAAAGACCACGTAAAAAGGAAGACAATGTCTTAGTTGAGAGCCATGAAAAATC 179

Qy     185 CC 186
           ||
Db     180 CC 181


RESULT 3
US-08-733-446-20
; Sequence 20, Application US/08733446
; Patent No. 5856138
;  GENERAL INFORMATION:
;    APPLICANT:  FUKUDA, Tsunehiko
;    TITLE OF INVENTION:  HUMAN PARATHYROID HORMONE MUTEINS AND
;    TITLE OF INVENTION:  PRODUCTION THEREOF
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
;      ADDRESSEE:  CUSHMAN
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  US
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/733,446
;      FILING DATE:  18-OCT-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/402,970
;      FILING DATE:
;      APPLICATION NUMBER:  US/07/926,787
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  NEUNER, George W
;      REGISTRATION NUMBER:  26964
;      REFERENCE/DOCKET NUMBER:  42025
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)523-3400
;      TELEFAX:  (617)523-6440
;      TELEX:  200291 STRE UR
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  237 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid, synthetic DNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..237
;      IDENTIFICATION METHOD:  E
US-08-733-446-20

  Query Match             8.9%;  Score 35.6;  DB 2;  Length 237;
  Best Local Similarity   55.7%;  Pred. No. 0.036;
  Matches   68;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;
```

```
Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
           ||| |||   | | |||||  |  |  |  || || ||   || |||  | ||   ||
Db      63 GAAGTTGCAGGATGTGCACAATTTTGTTGCCTTAGGTGCCCCATTGGCTCCTCGTGATGC 122

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
            | |||    |||  | | |||   || | ||| |||||    |||| | | |||||
Db     123 TGGTTCCCAAAGACCACGTAAAAAGGAAGACAATGTCTTAGTTGAGAGCCATGAAAAATC 182

Qy     185 CC 186
           ||
Db     183 CC 184


RESULT 4
US-08-733-446-21
; Sequence 21, Application US/08733446
; Patent No. 5856138
;  GENERAL INFORMATION:
;    APPLICANT:  FUKUDA, Tsunehiko
;    TITLE OF INVENTION:  HUMAN PARATHYROID HORMONE MUTEINS AND
;    TITLE OF INVENTION:  PRODUCTION THEREOF
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
;      ADDRESSEE:  CUSHMAN
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  US
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/733,446
;      FILING DATE:  18-OCT-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/402,970
;      FILING DATE:
;      APPLICATION NUMBER:  US/07/926,787
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  NEUNER, George W
;      REGISTRATION NUMBER:  26964
;      REFERENCE/DOCKET NUMBER:  42025
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)523-3400
;      TELEFAX:  (617)523-6440
;      TELEX:  200291 STRE UR
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  240 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid, synthetic DNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..240
;      IDENTIFICATION METHOD:  E
US-08-733-446-21

  Query Match             8.9%;  Score 35.6;  DB 2;  Length 240;
  Best Local Similarity   55.7%;  Pred. No. 0.036;
  Matches   68;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
           ||| |||   | | |||||  |  |  |  || || ||   || |||  | ||   ||
Db      66 GAAGTTGCAGGATGTGCACAATTTTGTTGCCTTAGGTGCCCCATTGGCTCCTCGTGATGC 125

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
            | |||    |||  | | |||   || | ||| |||||    |||| | | |||||
Db     126 TGGTTCCCAAAGACCACGTAAAAAGGAAGACAATGTCTTAGTTGAGAGCCATGAAAAATC 185

Qy     185 CC 186
           ||
Db     186 CC 187


RESULT 5
US-08-733-446-22
; Sequence 22, Application US/08733446
; Patent No. 5856138
;  GENERAL INFORMATION:
;    APPLICANT:  FUKUDA, Tsunehiko
;    TITLE OF INVENTION:  HUMAN PARATHYROID HORMONE MUTEINS AND
;    TITLE OF INVENTION:  PRODUCTION THEREOF
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
;      ADDRESSEE:  CUSHMAN
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  US
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/733,446
;      FILING DATE:  18-OCT-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/402,970
;      FILING DATE:
;      APPLICATION NUMBER:  US/07/926,787
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  NEUNER, George W
;      REGISTRATION NUMBER:  26964
;      REFERENCE/DOCKET NUMBER:  42025
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)523-3400
;      TELEFAX:  (617)523-6440
;      TELEX:  200291 STRE UR
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  243 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid, synthetic DNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..243
;      IDENTIFICATION METHOD:  E
US-08-733-446-22

  Query Match             8.9%;  Score 35.6;  DB 2;  Length 243;
  Best Local Similarity   55.7%;  Pred. No. 0.036;
  Matches   68;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
           ||| |||   | | |||||  |  |  |  || || ||   || |||  | ||   ||
Db      69 GAAGTTGCAGGATGTGCACAATTTTGTTGCCTTAGGTGCCCCATTGGCTCCTCGTGATGC 128

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
            | |||    |||  | | |||   || | ||| |||||    |||| | | |||||
Db     129 TGGTTCCCAAAGACCACGTAAAAAGGAAGACAATGTCTTAGTTGAGAGCCATGAAAAATC 188
```

```
Qy     185 CC 186
           ||
Db     189 CC 190


RESULT 6
US-08-733-446-24
; Sequence 24, Application US/08733446
; Patent No. 5856138
;  GENERAL INFORMATION:
;    APPLICANT:  FUKUDA, Tsunehiko
;    TITLE OF INVENTION:  HUMAN PARATHYROID HORMONE MUTEINS AND
;    TITLE OF INVENTION:  PRODUCTION THEREOF
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
;      ADDRESSEE:  CUSHMAN
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  US
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/733,446
;      FILING DATE:  18-OCT-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/402,970
;      FILING DATE:
;      APPLICATION NUMBER:  US/07/926,787
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  NEUNER, George W
;      REGISTRATION NUMBER:  26964
;      REFERENCE/DOCKET NUMBER:  42025
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)523-3400
;      TELEFAX:  (617)523-6440
;      TELEX:  200291 STRE UR
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  245 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid, synthetic DNA
;    ANTI-SENSE:  NO
US-08-733-446-24

  Query Match             8.9%;  Score 35.6;  DB 2;  Length 245;
  Best Local Similarity   55.7%;  Pred. No. 0.036;
  Matches   68;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
           ||| |||   | | |||||  |  |  |  || || ||   || |||  | ||   ||
Db      64 GAAGTTGCAGGATGTGCACAATTTTGTTGCCTTAGGTGCCCCATTGGCTCCTCGTGATGC 123

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
            | |||    |||  | | |||   || | ||| |||||    |||| | | |||||
Db     124 TGGTTCCCAAAGACCACGTAAAAAGGAAGACAATGTCTTAGTTGAGAGCCATGAAAAATC 183

Qy     185 CC 186
           ||
Db     184 CC 185


RESULT 7
US-08-733-446-25/c
; Sequence 25, Application US/08733446
; Patent No. 5856138
;  GENERAL INFORMATION:
;    APPLICANT:  FUKUDA, Tsunehiko
;    TITLE OF INVENTION:  HUMAN PARATHYROID HORMONE MUTEINS AND
;    TITLE OF INVENTION:  PRODUCTION THEREOF
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
;      ADDRESSEE:  CUSHMAN
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  US
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/733,446
;      FILING DATE:  18-OCT-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/402,970
;      FILING DATE:
;      APPLICATION NUMBER:  US/07/926,787
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  NEUNER, George W
;      REGISTRATION NUMBER:  26964
;      REFERENCE/DOCKET NUMBER:  42025
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)523-3400
;      TELEFAX:  (617)523-6440
;      TELEX:  200291 STRE UR
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  247 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid, synthetic DNA
;    ANTI-SENSE:  YES
US-08-733-446-25

  Query Match             8.9%;  Score 35.6;  DB 2;  Length 247;
  Best Local Similarity   55.7%;  Pred. No. 0.037;
  Matches   68;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
           ||| |||   | | |||||  |  |  |  || || ||   || |||  | ||   ||
Db     186 GAAGTTGCAGGATGTGCACAATTTTGTTGCCTTAGGTGCCCCATTGGCTCCTCGTGATGC 127

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
            | |||    |||  | | |||   || | ||| |||||    |||| | | |||||
Db     126 TGGTTCCCAAAGACCACGTAAAAAGGAAGACAATGTCTTAGTTGAGAGCCATGAAAAATC 67

Qy     185 CC 186
           ||
Db      66 CC 65


RESULT 8
US-08-733-446-26
; Sequence 26, Application US/08733446
; Patent No. 5856138
;  GENERAL INFORMATION:
;    APPLICANT:  FUKUDA, Tsunehiko
;    TITLE OF INVENTION:  HUMAN PARATHYROID HORMONE MUTEINS AND
;    TITLE OF INVENTION:  PRODUCTION THEREOF
;    NUMBER OF SEQUENCES:  62
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
;      ADDRESSEE:  CUSHMAN
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  US
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/733,446
;      FILING DATE:  18-OCT-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/402,970
;      FILING DATE:
;      APPLICATION NUMBER:  US/07/926,787
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  NEUNER, George W
;      REGISTRATION NUMBER:  26964
;      REFERENCE/DOCKET NUMBER:  42025
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)523-3400
;      TELEFAX:  (617)523-6440
;      TELEX:  200291 STRE UR
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  248 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid, synthetic DNA
;    ANTI-SENSE:  NO
US-08-733-446-26

  Query Match             8.9%;  Score 35.6;  DB 2;  Length 248;
  Best Local Similarity   55.7%;  Pred. No. 0.037;
  Matches   68;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
           ||| |||   | | ||||| |  |  |  || || ||   || |||  | ||   ||
Db     67 GAAGTTGCAGGATGTGCACAATTTTGTTGCCTTAGGTGCCCCATTGGCTCCTCGTGATGC 126

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
            | |||    |||  | | |||   || | ||| |||||    |||| | | |||||
Db     127 TGGTTCCCAAAGACCACGTAAAAAGGAAGACAATGTCTTAGTTGAGAGCCATGAAAAATC 186

Qy     185 CC 186
           ||
Db     187 CC 188


RESULT 9
US-08-733-446-27/c
; Sequence 27, Application US/08733446
; Patent No. 5856138
;  GENERAL INFORMATION:
;    APPLICANT:  FUKUDA, Tsunehiko
;    TITLE OF INVENTION:  HUMAN PARATHYROID HORMONE MUTEINS AND
;    TITLE OF INVENTION:  PRODUCTION THEREOF
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
;      ADDRESSEE:  CUSHMAN
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  US
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/733,446
;      FILING DATE:  18-OCT-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/402,970
;      FILING DATE:
;      APPLICATION NUMBER:  US/07/926,787
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  NEUNER, George W
;      REGISTRATION NUMBER:  26964
;      REFERENCE/DOCKET NUMBER:  42025
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)523-3400
;      TELEFAX:  (617)523-6440
;      TELEX:  200291 STRE UR
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  250 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid, synthetic DNA
;    ANTI-SENSE:  YES
US-08-733-446-27

  Query Match             8.9%;  Score 35.6;  DB 2;  Length 250;
  Best Local Similarity   55.7%;  Pred. No. 0.037;
  Matches   68;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
           ||| |||   | | ||||| |  |  |  || || ||   || |||  | ||   ||
Db     186 GAAGTTGCAGGATGTGCACAATTTTGTTGCCTTAGGTGCCCCATTGGCTCCTCGTGATGC 127

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
            | |||    |||  | | |||   || | ||| |||||    |||| | | |||||
Db     126 TGGTTCCCAAAGACCACGTAAAAAGGAAGACAATGTCTTAGTTGAGAGCCATGAAAAATC 67

Qy     185 CC 186
           ||
Db      66 CC 65


RESULT 10
US-08-733-446-28
; Sequence 28, Application US/08733446
; Patent No. 5856138
;  GENERAL INFORMATION:
;    APPLICANT:  FUKUDA, Tsunehiko
;    TITLE OF INVENTION:  HUMAN PARATHYROID HORMONE MUTEINS AND
;    TITLE OF INVENTION:  PRODUCTION THEREOF
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
;      ADDRESSEE:  CUSHMAN
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  US
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/08/733,446
;      FILING DATE:  18-OCT-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/402,970
;      FILING DATE:
;      APPLICATION NUMBER:  US/07/926,787
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  NEUNER, George W
;      REGISTRATION NUMBER:  26964
;      REFERENCE/DOCKET NUMBER:  42025
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)523-3400
;      TELEFAX:  (617)523-6440
;      TELEX:  200291 STRE UR
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  251 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid, synthetic DNA
;    ANTI-SENSE:  NO
US-08-733-446-28

  Query Match             8.9%;  Score 35.6;  DB 2;  Length 251;
  Best Local Similarity   55.7%;  Pred. No. 0.037;
  Matches   68;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
           ||| |||   | | |||||  |  |  |  || || ||   || |||  | ||   ||
Db      70 GAAGTTGCAGGATGTGCACAATTTTGTTGCCTTAGGTGCCCCATTGGCTCCTCGTGATGC 129

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
           | |||    |||  | | |||   || | ||| |||||    |||| | | |||||
Db     130 TGGTTCCCAAAGACCACGTAAAAAGGAAGACAATGTCTTAGTTGAGAGCCATGAAAAATC 189

Qy     185 CC 186
           ||
Db     190 CC 191


RESULT 11
US-08-689-190-1
; Sequence 1, Application US/08689190
; Patent No. 5714349
;  GENERAL INFORMATION:
;    APPLICANT:  FUKUDA, Tsunehiko
;    APPLICANT:  OSHIKA, Yuri
;    APPLICANT:  YAMADA, Takao
;    TITLE OF INVENTION:  SYNTHETIC GENE CODING FOR HUMAN
;    TITLE OF INVENTION:  PARATHYROID HORMONE
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
;      ADDRESSEE:  CUSHMAN
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  US
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/689,190
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/006,197
;      FILING DATE:
;      APPLICATION NUMBER:  US/08/016,171
;      FILING DATE:
;      APPLICATION NUMBER:  US/07/765,371
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  WILLIAMS, Gregory D
;      REGISTRATION NUMBER:  30,901
;      REFERENCE/DOCKET NUMBER:  41,288
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)523-3400
;      TELEFAX:  (617)523-6440
;      TELEX:  200291 STRE UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  252 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid, synthesizing DNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..252
;      IDENTIFICATION METHOD:  E
;    FEATURE:
;      NAME/KEY:  MUTATION
;      LOCATION:  7, 8, 9, 12, 15, 19, 21, 33, 36, 43, 51, 58, 60,
;      IDENTIFICATION METHOD:  S
US-08-689-190-1

  Query Match             8.9%;  Score 35.6;  DB 1;  Length 252;
  Best Local Similarity   55.7%;  Pred. No. 0.037;
  Matches   68;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
           ||| |||   | | |||||  |  |  |  || || ||   || |||  | ||   ||
Db      78 GAAGTTGCAGGATGTGCACAATTTTGTTGCCTTAGGTGCCCCATTGGCTCCTCGTGATGC 137

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
           | |||  |||  | | |||   || | ||| |||||    |||| | | |||||
Db     138 TGGTTCCCAAAGACCACGTAAAAAGGAAGACAATGTCTTAGTTGAGAGCCATGAAAAATC 197

Qy     185 CC 186
           ||
Db     198 CC 199


RESULT 12
US-08-733-446-23
; Sequence 23, Application US/08733446
; Patent No. 5856138
;  GENERAL INFORMATION:
;    APPLICANT:  FUKUDA, Tsunehiko
;    TITLE OF INVENTION:  HUMAN PARATHYROID HORMONE MUTEINS AND
;    TITLE OF INVENTION:  PRODUCTION THEREOF
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
;      ADDRESSEE:  CUSHMAN
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  US
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/733,446
;      FILING DATE:  18-OCT-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/08/402,970
;      FILING DATE:
;      APPLICATION NUMBER:  US/07/926,787
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  NEUNER, George W
;      REGISTRATION NUMBER:  26964
;      REFERENCE/DOCKET NUMBER:  42025
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)523-3400
;      TELEFAX:  (617)523-6440
;      TELEX:  200291 STRE UR
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  252 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid,  synthetic DNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..252
;      IDENTIFICATION METHOD:  E
US-08-733-446-23

  Query Match             8.9%;  Score 35.6;  DB 2;  Length 252;
  Best Local Similarity   55.7%;  Pred. No. 0.037;
  Matches   68;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
           ||| |||   | | |||||  |  |  |  || || ||   || |||  | ||   ||
Db      78 GAAGTTGCAGGATGTGCACAATTTTGTTGCCTTAGGTGCCCCATTGGCTCCTCGTGATGC 137

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
            | |||    |||  | | |||   || | ||| |||||    |||| | | |||||
Db     138 TGGTTCCCAAAGACCACGTAAAAAGGAAGACAATGTCTTAGTTGAGAGCCATGAAAAATC 197

Qy     185 CC 186
           ||
Db     198 CC 199


RESULT 13
US-08-733-446-56
; Sequence 56, Application US/08733446
; Patent No. 5856138
;  GENERAL INFORMATION:
;    APPLICANT:  FUKUDA, Tsunehiko
;    TITLE OF INVENTION:  HUMAN PARATHYROID HORMONE MUTEINS AND
;    TITLE OF INVENTION:  PRODUCTION THEREOF
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
;      ADDRESSEE:  CUSHMAN
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  US
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/733,446
;      FILING DATE:  18-OCT-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/402,970
;      FILING DATE:
;      APPLICATION NUMBER:  US/07/926,787
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  NEUNER, George W
;      REGISTRATION NUMBER:  26964
;      REFERENCE/DOCKET NUMBER:  42025
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)523-3400
;      TELEFAX:  (617)523-6440
;      TELEX:  200291 STRE UR
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  252 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid, synthetic DNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..252
;      NAME/KEY:  mutation
;      LOCATION:  103...105
;      IDENTIFICATION METHOD:  E
US-08-733-446-56

  Query Match             8.9%;  Score 35.6;  DB 2;  Length 252;
  Best Local Similarity   55.7%;  Pred. No. 0.037;
  Matches   68;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
           ||| |||   | | |||||  |  |  |  || || ||   || |||  | ||   ||
Db      78 GAAGTTGCAGGATGTGCACAATTTTTGCGCCTTAGGTGCCCCATTGGCTCCTCGTGATGC 137

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
            | |||    |||  | | |||   || | ||| |||||    |||| | | |||||
Db     138 TGGTTCCCAAAGACCACGTAAAAAGGAAGACAATGTCTTAGTTGAGAGCCATGAAAAATC 197

Qy     185 CC 186
           ||
Db     198 CC 199


RESULT 14
US-08-733-446-57
; Sequence 57, Application US/08733446
; Patent No. 5856138
;  GENERAL INFORMATION:
;    APPLICANT:  FUKUDA, Tsunehiko
;    TITLE OF INVENTION:  HUMAN PARATHYROID HORMONE MUTEINS AND
;    TITLE OF INVENTION:  PRODUCTION THEREOF
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
;      ADDRESSEE:  CUSHMAN
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  US
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/733,446
;      FILING DATE:  18-OCT-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/402,970
;      FILING DATE:
;      APPLICATION NUMBER:  US/07/926,787
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  NEUNER, George W
```

```
;       REGISTRATION NUMBER:  26964
;       REFERENCE/DOCKET NUMBER:  42025
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617)523-3400
;       TELEFAX:  (617)523-6440
;       TELEX:  200291 STRE UR
;   INFORMATION FOR SEQ ID NO:  57:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  252 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  other nucleic acid, synthetic DNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..252
;       NAME/KEY:  mutation
;       LOCATION:  51...52
;       IDENTIFICATION METHOD:  E
US-08-733-446-57

  Query Match              8.9%;  Score 35.6;  DB 2;  Length 252;
  Best Local Similarity   55.7%;  Pred. No. 0.037;
  Matches   68;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
           ||| |||   | | |||||  |  |  |  || || ||   || |||  | ||   ||
Db      78 GAAGTTGCAGGATGTGCACAATTTTGTTGCCTTAGGTGCCCCATTGGCTCCTCGTGATGC 137

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
           | |||    |||  | | |||   || | ||| |||||    |||| | | |||||
Db     138 TGGTTCCCAAAGACCACGTAAAAAGGAAGACAATGTCTTAGTTGAGAGCCATGAAAAATC 197

Qy     185 CC 186
           ||
Db     198 CC 199


RESULT 15
US-08-733-446-58
; Sequence 58, Application US/08733446
; Patent No. 5856138
;  GENERAL INFORMATION:
;    APPLICANT:  FUKUDA, Tsunehiko
;    TITLE OF INVENTION:  HUMAN PARATHYROID HORMONE MUTEINS AND
;    TITLE OF INVENTION:  PRODUCTION THEREOF
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
;      ADDRESSEE:  CUSHMAN
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  US
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/733,446
;      FILING DATE:  18-OCT-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/402,970
;      FILING DATE:
;      APPLICATION NUMBER:  US/07/926,787
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  NEUNER, George W
;      REGISTRATION NUMBER:  26964
;      REFERENCE/DOCKET NUMBER:  42025
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)523-3400
;      TELEFAX:  (617)523-6440
;      TELEX:  200291 STRE UR
;  INFORMATION FOR SEQ ID NO:  58:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  252 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid, synthetic DNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..252
;      NAME/KEY:  mutation
;      LOCATION:  19...24
;      IDENTIFICATION METHOD:  E
US-08-733-446-58

  Query Match              8.9%;  Score 35.6;  DB 2;  Length 252;
  Best Local Similarity   55.7%;  Pred. No. 0.037;
  Matches   68;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
           ||| |||   | | |||||  |  |  |  || || ||   || |||  | ||   ||
Db      78 GAAGTTGCAGGATGTGCACAATTTTGTTGCCTTAGGTGCCCCATTGGCTCCTCGTGATGC 137

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
           | |||    |||  | | |||   || | ||| |||||    |||| | | |||||
Db     138 TGGTTCCCAAAGACCACGTAAAAAGGAAGACAATGTCTTAGTTGAGAGCCATGAAAAATC 197

Qy     185 CC 186
           ||
Db     198 CC 199


Search completed: January  2, 2003, 13:22:05
Job time : 76 secs
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: January 2, 2003, 13:22:11 ; Search time 84 Seconds
(without alignments)
2069.842 Million cell updates/sec

Title: US-09-850-147-1
Perfect score: 402
Sequence: 1 ccgggccgggctcaccagct.....ggcataacttcctgtgttcc 402

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 381593 seqs, 216252194 residues

Total number of hits satisfying chosen parameters: 763186

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	36.4	9.1	2553	10	US-09-835-995-1	Sequence 1, Appli
c 2	34.8	8.7	1984	9	US-09-854-133-390	Sequence 390, App
c 3	34.8	8.7	1984	10	US-09-738-973-390	Sequence 390, App
c 4	33.4	8.3	989	10	US-09-070-927A-618	Sequence 618, App
5	32.2	8.0	642	10	US-09-764-847-1449	Sequence 1449, Ap
6	32.2	8.0	642	10	US-09-764-847-1450	Sequence 1450, Ap
7	32.2	8.0	642	10	US-09-764-847-1451	Sequence 1451, Ap
8	32.2	8.0	642	10	US-09-764-847-1452	Sequence 1452, Ap
c 9	32.2	8.0	855	10	US-09-815-242-4500	Sequence 4500, Ap
c 10	32.2	8.0	861	10	US-09-815-242-8275	Sequence 8275, Ap
11	32	8.0	180216	10	US-09-835-232-6	Sequence 6, Appli
12	31.6	7.9	1984	10	US-09-864-761-4541	Sequence 4541, Ap
c 13	31.2	7.8	401	9	US-09-946-807-423	Sequence 423, App
c 14	31.2	7.8	401	9	US-09-946-807-424	Sequence 424, App
c 15	31.2	7.8	401	9	US-09-946-807-425	Sequence 425, App
c 16	31.2	7.8	401	9	US-09-946-807-426	Sequence 426, App
c 17	31.2	7.8	401	10	US-09-795-668-423	Sequence 423, App
c 18	31.2	7.8	401	10	US-09-795-668-424	Sequence 424, App
c 19	31.2	7.8	401	10	US-09-795-668-425	Sequence 425, App
c 20	31.2	7.8	401	10	US-09-795-668-426	Sequence 426, App
c 21	31.2	7.8	401	10	US-09-795-686-423	Sequence 423, App
c 22	31.2	7.8	401	10	US-09-795-686-424	Sequence 424, App
c 23	31.2	7.8	401	10	US-09-795-686-425	Sequence 425, App
c 24	31.2	7.8	401	10	US-09-795-686-426	Sequence 426, App
c 25	31.2	7.8	643	10	US-09-764-847-264	Sequence 264, App
c 26	31	7.7	411	10	US-09-983-965-4361	Sequence 4361, Ap
27	31	7.7	29793	10	US-09-973-451-38	Sequence 38, Appl
c 28	30.8	7.7	434	10	US-09-960-352-12123	Sequence 12123, A
29	30.8	7.7	2141	10	US-09-841-786-11	Sequence 11, Appl
30	30.8	7.7	9726	10	US-09-841-786-8	Sequence 8, Appli
31	30.8	7.7	11130	10	US-09-841-786-15	Sequence 15, Appl
c 32	30.6	7.6	861	10	US-09-815-242-8879	Sequence 8879, Ap
c 33	30.6	7.6	861	10	US-09-815-242-9010	Sequence 9010, Ap
34	30.6	7.6	2594	10	US-09-925-300-276	Sequence 276, App
c 35	30.6	7.6	3030	10	US-09-906-453-3	Sequence 3, Appli
36	30.6	7.6	1503841	9	US-09-946-807-1	Sequence 1, Appli
37	30.6	7.6	1503841	10	US-09-795-668-1	Sequence 1, Appli
38	30.6	7.6	1503841	10	US-09-795-686-1	Sequence 1, Appli
c 39	30.4	7.6	388	10	US-09-983-965-3462	Sequence 3462, Ap
c 40	30.4	7.6	2606	9	US-10-153-273-7	Sequence 7, Appli
41	30.4	7.6	2660	10	US-09-833-381-1015	Sequence 1015, Ap
42	30.2	7.5	355	10	US-09-960-352-7539	Sequence 7539, Ap
c 43	30.2	7.5	722	9	US-09-938-842A-3151	Sequence 3151, Ap
c 44	30	7.5	374	10	US-09-895-828-286	Sequence 286, App
c 45	29.8	7.4	2000	9	US-09-938-842A-3362	Sequence 3362, Ap

ALIGNMENTS

RESULT 1
US-09-835-995-1/c
; Sequence 1, Application US/09835995
; Patent No. US20020052021A1
; GENERAL INFORMATION:
; APPLICANT: Soderberg, Charlotte
; APPLICANT: Lind, Peter
; TITLE OF INVENTION: No. US20020052021A1el G Protein Coupled Receptors
; FILE REFERENCE: 00146reg.US
; CURRENT APPLICATION NUMBER: US/09/835,995
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 60/198,090
; PRIOR FILING DATE: 2000-04-17
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2553
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-835-995-1

Query Match 9.1%; Score 36.4; DB 10; Length 2553;
Best Local Similarity 54.5%; Pred. No. 0.15;
Matches 73; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

```
Qy    151 ATGTCAAGGTCTTCACAGAGACCGAGGAAAACAACCAGATTCCAGGGCCAAGGGCTTGGA 210
          ||  |||| |||  |     || || |||      ||||  ||| | | |  | || | |
Db    828 ATAACAAGCTCTGGAAGAGTACAGAAGAATGACTACAGAGCCCACGACGACAGTCTGGCA 769

Qy    211 ATCCCATATTTCCTTTGACTTCAGGTGGACAATTTTAGAAGGCAATATTTAAATCTGGAA 270
              | ||  ||||||| |  |||| |||| |    || | || | ||  ||| | | |
Db    768 CAGAGACATACCCTTTGATTCGAGGTAGACATTAGCTGATGACATTTTTGTAATTTTGCA 709

Qy    271 GGGGTAAAGAACAA 284
            ||||   || |
Db    708 TATGTAACACACTA 695
```

RESULT 2
US-09-854-133-390/c
; Sequence 390, Application US/09854133
; Publication No. US20020183499A1

```
; GENERAL INFORMATION:
;  APPLICANT: Lodes, Michael J.
;  APPLICANT:  Mohamath, Raodoh
;  APPLICANT:  Henderson, Robert A.
;  APPLICANT:  Benson, Darin R.
;  APPLICANT:  Secrist, Heather
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
;  TITLE OF INVENTION:  THE THERAPY AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.475C10
;  CURRENT APPLICATION NUMBER: US/09/854,133
;  CURRENT FILING DATE:  2001-05-11
;  NUMBER OF SEQ ID NOS: 735
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 390
;   LENGTH: 1984
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-854-133-390

  Query Match             8.7%;  Score 34.8;  DB 9;  Length 1984;
  Best Local Similarity   53.7%;  Pred. No. 0.41;
  Matches   72;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy    151 ATGTCAAGGTCTTCACAGAGACCGAGGAAAACAACCAGATTCCAGGGCCAAGGGCTTGGA 210
          ||  |||| |||  |     || || |||      ||||  ||| |   |  | || | |
Db    908 ATAACAAGCTCTGGAAGAGTACAGAAGAATGACTACAGAGCCCACGATGACAGTCTGGCA 849

Qy    211 ATCCCATATTTCCTTTGACTTCAGGTGGACAATTTTAGAAGGCAATATTTAAATCTGGAA 270
            | ||  |||||||  |  |||| |||| |    || | || | ||  ||| | | |
Db    848 CAGAGACATACCCTTTGATTCGAGGTAGACATTAGCTGATGACATTTTTGTAATTTTGCA 789

Qy    271 GGGGTAAAGAACAA 284
            ||||   || |
Db    788 TATGTAACACACTA 775


RESULT 3
US-09-738-973-390/c
; Sequence 390, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Henderson, Robert A.
;  APPLICANT:  Lodes, Michael J.
;  APPLICANT:  Fling, Steven P.
;  APPLICANT:  Mohamath, Raodoh
;  APPLICANT:  Algate, Paul A.
;  APPLICANT:  Secrist, Heather
;  APPLICANT:  Indirias, Carol Yoseph
;  APPLICANT:  Benson, Darin R.
;  APPLICANT:  Elliot, Mark
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Kalos, Michael D.
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR
;  TITLE OF INVENTION:  THE THERAPY AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.475C9
;  CURRENT APPLICATION NUMBER: US/09/738,973
;  CURRENT FILING DATE:  2000-12-14
;  NUMBER OF SEQ ID NOS:  587
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 390
;   LENGTH: 1984
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-738-973-390

  Query Match             8.7%;  Score 34.8;  DB 10;  Length 1984;
  Best Local Similarity   53.7%;  Pred. No. 0.41;
  Matches   72;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy    151 ATGTCAAGGTCTTCACAGAGACCGAGGAAAACAACCAGATTCCAGGGCCAAGGGCTTGGA 210
          ||  |||| |||  |     || || |||      ||||  ||| |   |  | || | |
Db    908 ATAACAAGCTCTGGAAGAGTACAGAAGAATGACTACAGAGCCCACGATGACAGTCTGGCA 849

Qy    211 ATCCCATATTTCCTTTGACTTCAGGTGGACAATTTTAGAAGGCAATATTTAAATCTGGAA 270
            | ||  |||||||  |  |||| |||| |    || | || | ||  ||| | | |
Db    848 CAGAGACATACCCTTTGATTCGAGGTAGACATTAGCTGATGACATTTTTGTAATTTTGCA 789

Qy    271 GGGGTAAAGAACAA 284
            ||||   || |
Db    788 TATGTAACACACTA 775


RESULT 4
US-09-070-927A-618/c
; Sequence 618, Application US/09070927A
; Patent No. US20020120116A1
;   GENERAL INFORMATION:
;        APPLICANT: Charles A. Kunsch
;                   Patrick J. Dillon
;                   Steven Barash
;        TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
;        NUMBER OF SEQUENCES: 982
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Human Genome Sciences, Inc.
;             STREET: 9410 Key West Avenue
;             CITY: Rockville
;             STATE: Maryland
;             COUNTRY: USA
;             ZIP: 20850
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;             COMPUTER: HP Vectra 486/33
;             OPERATING SYSTEM: MSDOS version 6.2
;             SOFTWARE: ASCII Text
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/070,927A
;             FILING DATE: 04-May-2000
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/046,655
;             FILING DATE: 1997-05-16
;             APPLICATION NUMBER: 60/044,031
;             FILING DATE: 1997-05-06
;             APPLICATION NUMBER: 60/066,009
;             FILING DATE: 1997-11-14
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Kenley K. Hoover
;             REGISTRATION NUMBER: 40,302
;             REFERENCE/DOCKET NUMBER: PB369
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (301) 309-8504
;             TELEFAX: (301) 309-8512
;   INFORMATION FOR SEQ ID NO: 618:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 989 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 618:
US-09-070-927A-618

  Query Match             8.3%;  Score 33.4;  DB 10;  Length 989;
  Best Local Similarity   62.7%;  Pred. No. 0.79;
  Matches   52;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy    238 GACAATTTTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTA 297
          || ||| | || ||||||| |||||   ||  |  | ||||  ||    || |  || ||
Db    655 GAAAATATAAGGAGGCAATTTTTAATATTGCCAATGATAAATTACTTATAGCTTTTGGTA 596

Qy    298 ATTGTGGAGGTTCAGATCATAAT 320
            || ||| |||  |||  | |||
Db    595 TTTTTGGTGGTATAGAAGAAAAT 573


RESULT 5
```

```
US-09-764-847-1449
; Sequence 1449, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC009
;  CURRENT APPLICATION NUMBER: US/09/764,847
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 2003
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1449
;   LENGTH: 642
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-847-1449

  Query Match             8.0%;  Score 32.2;  DB 10;  Length 642;
  Best Local Similarity   63.6%;  Pred. No. 1.5;
  Matches   49;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy     277 AAGAACAATCAGATGGTGTTAATTGTGGAGGTTCAGATCATAATGCCATCTGATGCATTC 336
           || ||   ||| |||||| ||| ||||| ||   | |  | ||| |  ||| ||| |||
Db     311 AAAAATTGTCACATGGTGGTAAGTGTGGTGGAGAAAAATAAAATACACTCTAATGAATTT 370

Qy     337 CATTTCTGGTTTCTTTT 353
           ||||| |  | |  | |
Db     371 CATTTTTTTCTTCTGT 387


RESULT 6
US-09-764-847-1450
; Sequence 1450, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC009
;  CURRENT APPLICATION NUMBER: US/09/764,847
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 2003
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1450
;   LENGTH: 642
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-847-1450

  Query Match             8.0%;  Score 32.2;  DB 10;  Length 642;
  Best Local Similarity   63.6%;  Pred. No. 1.5;
  Matches   49;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy     277 AAGAACAATCAGATGGTGTTAATTGTGGAGGTTCAGATCATAATGCCATCTGATGCATTC 336
           || ||   ||| |||||| ||| ||||| ||   | |  | ||| |  ||| ||| |||
Db     311 AAAAATTGTCACATGGTGGTAAGTGTGGTGGAGAAAAATAAAATACACTCTAATGAATTT 370

Qy     337 CATTTCTGGTTTCTTTT 353
           ||||| |  | |  | |
Db     371 CATTTTTTTCTTCTGT 387


RESULT 7
US-09-764-847-1451
; Sequence 1451, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC009
;  CURRENT APPLICATION NUMBER: US/09/764,847
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 2003
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1451
;   LENGTH: 642
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-847-1451

  Query Match             8.0%;  Score 32.2;  DB 10;  Length 642;
  Best Local Similarity   63.6%;  Pred. No. 1.5;
  Matches   49;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy     277 AAGAACAATCAGATGGTGTTAATTGTGGAGGTTCAGATCATAATGCCATCTGATGCATTC 336
           || ||   ||| |||||| ||| ||||| ||   | |  | ||| |  ||| ||| |||
Db     311 AAAAATTGTCACATGGTGGTAAGTGTGGTGGAGAAAAATAAAATACACTCTAATGAATTT 370

Qy     337 CATTTCTGGTTTCTTTT 353
           ||||| |  | |  | |
Db     371 CATTTTTTTCTTCTGT 387


RESULT 8
US-09-764-847-1452
; Sequence 1452, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC009
;  CURRENT APPLICATION NUMBER: US/09/764,847
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 2003
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1452
;   LENGTH: 642
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-847-1452

  Query Match             8.0%;  Score 32.2;  DB 10;  Length 642;
  Best Local Similarity   63.6%;  Pred. No. 1.5;
  Matches   49;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy     277 AAGAACAATCAGATGGTGTTAATTGTGGAGGTTCAGATCATAATGCCATCTGATGCATTC 336
           || ||   ||| |||||| ||| ||||| ||   | |  | ||| |  ||| ||| |||
Db     311 AAAAATTGTCACATGGTGGTAAGTGTGGTGGAGAAAAATAAAATACACTCTAATGAATTT 370

Qy     337 CATTTCTGGTTTCTTTT 353
           ||||| |  | |  | |
Db     371 CATTTTTTTCTTCTGT 387


RESULT 9
US-09-815-242-4500/c
; Sequence 4500, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
```

```
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4500
;   LENGTH: 855
;   TYPE: DNA
;   ORGANISM: Staphylococcus aureus
US-09-815-242-4500

  Query Match             8.0%;  Score 32.2;  DB 10;  Length 855;
  Best Local Similarity   51.0%;  Pred. No. 1.8;
  Matches   76;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy    206 TTGGAATCCCATATTTCCTTTGACTTCAGGTGGACAATTTTAGAAGGCAATATTTAAATC 265
          ||||  | |  ||||| | || | || || || || || |  |   |  ||  | |  | |||
Db    711 TTGGTTTTCAGTATTTACGTTAATTTTAGCTGTACCAAATGGAATTGCTTTTTGGATATC 652

Qy    266 TGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGGAGGTTCAGATCATAATGCCAT 325
          |  |   || | |  |  | ||    ||  || |  |||     | |   |||| | ||
Db    651 TTTAGTCGGGATACCAGTACCACCGTGTAATACTAATGGTAAACCTGTAGATAAACCGAT 592

Qy    326 CTGATGCATTCCATTTCTGGTTTCTTTTG 354
            |  | |||| | ||      |  |||||
Db    591 TTCTTCCATTTCTTTAAATCCTAATTTTG 563


RESULT 10
US-09-815-242-8275/c
; Sequence 8275, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8275
;   LENGTH: 861
;   TYPE: DNA
;   ORGANISM: Staphylococcus aureus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(861)
US-09-815-242-8275

  Query Match             8.0%;  Score 32.2;  DB 10;  Length 861;
  Best Local Similarity   51.0%;  Pred. No. 1.8;
  Matches   76;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy    206 TTGGAATCCCATATTTCCTTTGACTTCAGGTGGACAATTTTAGAAGGCAATATTTAAATC 265
          ||||  | |  ||||| | || | || || || || || |  |   |  ||  | |  | |||
Db    711 TTGGTTTTCAGTATTTACGTTAATTTTAGCTGTACCAAATGGAATTGCTTTTTGGATATC 652

Qy    266 TGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGGAGGTTCAGATCATAATGCCAT 325
          |  |   || | |  |  | ||    ||  || |  |||     | |   |||| | ||
Db    651 TTTAGTCGGGATACCAGTACCACCGTGTAATACTAATGGTAAACCTGTAGATAAACCGAT 592

Qy    326 CTGATGCATTCCATTTCTGGTTTCTTTTG 354
            |  | |||| | ||      |  |||||
Db    591 TTCTTCCATTTCTTTAAATCCTAATTTTG 563


RESULT 11
US-09-835-232-6
; Sequence 6, Application US/09835232
; Patent No. US20020098489A1
; GENERAL INFORMATION:
;  APPLICANT: Leder, Philip
;  APPLICANT:  Leader, Benjamin
;  TITLE OF INVENTION: FORMIN-2 NUCLEIC ACIDS AND POLYPEPTIDES
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: 00383/052002
;  CURRENT APPLICATION NUMBER: US/09/835,232
;  CURRENT FILING DATE:  2001-04-12
;  PRIOR APPLICATION NUMBER: US 60/196,811
;  PRIOR FILING DATE: 2000-04-13
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
;   LENGTH: 180216
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(180216)
;   OTHER INFORMATION: n = A,T,C or G
US-09-835-232-6

  Query Match             8.0%;  Score 32;  DB 10;  Length 180216;
  Best Local Similarity   51.4%;  Pred. No. 38;
  Matches   74;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy    131 GATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAACCAGAT 190
          ||  | || ||  |||| ||| |   | |          | |  |   |||||||||| ||
Db  22235 GAAAAAAAAAAAAAAAGAAGAAGAAGAAGGGGGGAAAAAAGGACAAAACAACCATAT 22294

Qy    191 TCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGACAATTTTAGAA 250
          ||||      |  | |||  |  | |   |   | | | ||| |   |||||  | || |
Db  22295 TCCATAAAGCATTGATTGTTACACAAATATGGTTCTCAGTTCTGTATGACAAATATAAAG 22354

Qy    251 GGCAATATTTAAATCTGGAAGGGG 274
            |   | || ||| | ||   |||
Db  22355 TGTTTTCTTAAAACCAGGCTTGGG 22378


RESULT 12
US-09-864-761-4541
; Sequence 4541, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 4541
;   LENGTH: 1984
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC002094.1
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  31
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.9
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL =  31
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 7.6
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL =  36
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL =  11
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 8.8
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9
US-09-864-761-4541

  Query Match             7.9%;  Score 31.6;  DB 10;  Length 1984;
  Best Local Similarity   48.4%;  Pred. No. 4.4;
  Matches   88;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy     207 TGGAATCCCATATTTCCTTTGACTTCAGGTGGACAATTTTAGAAGGCAATATTTAAATCT 266
           | ||||    ||||| |   |  |    |   |   ||| |||     |||||     |
Db    1712 TTGAATTTTCTATTTATTAGTATATACATTTTTCTTCTTTTGAAATGTGTATTTTCTAGT 1771

Qy     267 GGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGGAGGTTCAGATCATAATGCCATC 326
           |  || ||  ||| |       | |  |||    || |  || || |||   |  | |
Db    1772 TTATTGGATATTGAAAACAATCTTTGATTTATCATTGTAATTTAAGTTCACTGTCTCTTT 1831

Qy     327 TGATGCATTCCATTTCTGGTTTCTTTTGAACACAAGTACAAAAAACCAACGGGAGGGGCA 386
           |   ||| |   | | |  | |||||||  ||  |||| |||| |  | |       |||
Db    1832 TCCAGCAATTATTCTTTTTTCTCTTTTGTGCAGTAGTAGAAAACAAAATCCAAGCAAGCA 1891

Qy     387 TA 388
           ||
Db    1892 TA 1893


RESULT 13
US-09-946-807-423/c
; Sequence 423, Application US/09946807
; Patent No. US20020165144A1
; GENERAL INFORMATION:
;  APPLICANT: Stefansson, Hreinn
;  APPLICANT:  Steinthorsdottir, Valgerdur
;  APPLICANT:  Gulcher, Jeffrey R.
;  TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
;  FILE REFERENCE: 2345.2004-001
;  CURRENT APPLICATION NUMBER: US/09/946,807
;  CURRENT FILING DATE:  2001-09-05
;  PRIOR APPLICATION NUMBER: US/09/795,668
;  PRIOR FILING DATE: 2001-02-28
;  PRIOR APPLICATION NUMBER: US 09/515,716
;  PRIOR FILING DATE: 2000-02-28
;  NUMBER OF SEQ ID NOS: 1531
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 423
;   LENGTH: 401
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-946-807-423

  Query Match             7.8%;  Score 31.2;  DB 9;  Length 401;
  Best Local Similarity   47.6%;  Pred. No. 2.4;
  Matches   81;  Conservative    3;  Mismatches   86;  Indels    0;  Gaps    0;

Qy     178 AAAACAACCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTG 237
           |:|| |   |||  |      | |     || |   ||   |||||| |  |   |||
Db     301 ARAAAATTGAGAGGCAGCTTCTAGTTCTTTCGGGCCCTTCATTTCCTATTTCCCAAGGRA 242

Qy     238 GACAATTTTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTA 297
           | |  ||| |  |      |  :||| |||  ||  |  :  ||||| |||| ||| ||
Db     241 AGCCAAGTTAAAGTGTCTACTCARAATGTGGCCGGCATTTWACACAATTAGATTGTGATA 182

Qy     298 ATTGTGGAGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTT 347
           |     ||  || | |    | || | | | |  |||   | || |  ||
Db     181 ACAAGTGAATTTGACAAAGCATTGTCTTTTTAATCATGGTAGTTTTCCTT 132


RESULT 14
US-09-946-807-424/c
; Sequence 424, Application US/09946807
; Patent No. US20020165144A1
; GENERAL INFORMATION:
;  APPLICANT: Stefansson, Hreinn
;  APPLICANT:  Steinthorsdottir, Valgerdur
;  APPLICANT:  Gulcher, Jeffrey R.
;  TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
;  FILE REFERENCE: 2345.2004-001
;  CURRENT APPLICATION NUMBER: US/09/946,807
;  CURRENT FILING DATE:  2001-09-05
;  PRIOR APPLICATION NUMBER: US/09/795,668
;  PRIOR FILING DATE: 2001-02-28
;  PRIOR APPLICATION NUMBER: US 09/515,716
;  PRIOR FILING DATE: 2000-02-28
;  NUMBER OF SEQ ID NOS: 1531
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 424
```

```
;   LENGTH: 401
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-946-807-424

  Query Match             7.8%;  Score 31.2;  DB 9;  Length 401;
  Best Local Similarity   47.6%;  Pred. No. 2.4;
  Matches   81;  Conservative    3;  Mismatches   86;  Indels    0;  Gaps    0;

Qy    178 AAAACAACCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTG 237
          |:|| |   |||  |     | |     || |   ||   ||||||| |  |   |||
Db    284 ARAAAATTGAGAGGCAGCTTCTAGTTCTTTCGGGCCCTTCATTTCCTATTTCCCAAGGRA 225

Qy    238 GACAATTTTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTA 297
            | |  ||| |  |     |  :||| |||  ||  |  :  ||||| |||| ||| ||
Db    224 AGCCAAGTTAAAGTGTCTACTCARAATGTGGCCGGCATTTWACACAATTAGATTGTGATA 165

Qy    298 ATTGTGGAGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTT 347
          |     ||  || | |    | || | | | |  |||   | || |  ||
Db    164 ACAAGTGAATTTGACAAAGCATTGTCTTTTTAATCATGGTAGTTTTCCTT 115


RESULT 15
US-09-946-807-425/c
; Sequence 425, Application US/09946807
; Patent No. US20020165144A1
; GENERAL INFORMATION:
;  APPLICANT: Stefansson, Hreinn
;  APPLICANT:  Steinthorsdottir, Valgerdur
;  APPLICANT:  Gulcher, Jeffrey R.
;  TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
;  FILE REFERENCE: 2345.2004-001
;  CURRENT APPLICATION NUMBER: US/09/946,807
;  CURRENT FILING DATE:  2001-09-05
;  PRIOR APPLICATION NUMBER: US/09/795,668
;  PRIOR FILING DATE: 2001-02-28
;  PRIOR APPLICATION NUMBER: US 09/515,716
;  PRIOR FILING DATE: 2000-02-28
;  NUMBER OF SEQ ID NOS: 1531
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 425
;   LENGTH: 401
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-946-807-425

  Query Match             7.8%;  Score 31.2;  DB 9;  Length 401;
  Best Local Similarity   47.6%;  Pred. No. 2.4;
  Matches   81;  Conservative    3;  Mismatches   86;  Indels    0;  Gaps    0;

Qy    178 AAAACAACCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTG 237
          |:|| |   |||  |     | |     || |   ||   ||||||| |  |   |||
Db    259 ARAAAATTGAGAGGCAGCTTCTAGTTCTTTCGGGCCCTTCATTTCCTATTTCCCAAGGRA 200

Qy    238 GACAATTTTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTA 297
            | |  ||| |  |     |  :||| |||  ||  |  :  ||||| |||| ||| ||
Db    199 AGCCAAGTTAAAGTGTCTACTCARAATGTGGCCGGCATTTWACACAATTAGATTGTGATA 140

Qy    298 ATTGTGGAGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTT 347
          |     ||  || | |    | || | | | |  |||   | || |  ||
Db    139 ACAAGTGAATTTGACAAAGCATTGTCTTTTTAATCATGGTAGTTTTCCTT 90


Search completed: January  2, 2003, 15:06:42
Job time : 129 secs
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: January 2, 2003, 12:27:57 ; Search time 3761 Seconds
(without alignments)
2687.392 Million cell updates/sec

Title: US-09-850-147-1
Perfect score: 402
Sequence: 1 ccgggccgggctcaccagct.....ggcataacttcctgtgttcc 402

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 24791104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/1/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/1/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/1/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/1/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/1/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/1/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/1/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/1/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/1/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/1/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/1/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/1/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/1/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/1/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/1/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/1/pna/US092_COMB.seq:*
17: /cgn2_6/ptodata/1/pna/US093_COMB.seq:*
18: /cgn2_6/ptodata/1/pna/US094_COMB.seq:*
19: /cgn2_6/ptodata/1/pna/US095A_COMB.seq:*
20: /cgn2_6/ptodata/1/pna/US095B_COMB.seq:*
21: /cgn2_6/ptodata/1/pna/US095C_COMB.seq:*
22: /cgn2_6/ptodata/1/pna/US095D_COMB.seq:*
23: /cgn2_6/ptodata/1/pna/US096A_COMB.seq:*
24: /cgn2_6/ptodata/1/pna/US096B_COMB.seq:*
25: /cgn2_6/ptodata/1/pna/US096C_COMB.seq:*
26: /cgn2_6/ptodata/1/pna/US096D_COMB.seq:*
27: /cgn2_6/ptodata/1/pna/US096E_COMB.seq:*
28: /cgn2_6/ptodata/1/pna/US097A_COMB.seq:*
29: /cgn2_6/ptodata/1/pna/US097B_COMB.seq:*
30: /cgn2_6/ptodata/1/pna/US097C_COMB.seq:*
31: /cgn2_6/ptodata/1/pna/US098A_COMB.seq:*
32: /cgn2_6/ptodata/1/pna/US098B_COMB.seq:*
33: /cgn2_6/ptodata/1/pna/US098C_COMB.seq:*
34: /cgn2_6/ptodata/1/pna/US099A_COMB.seq:*
35: /cgn2_6/ptodata/1/pna/US099B_COMB.seq:*
36: /cgn2_6/ptodata/1/pna/US099C_COMB.seq:*
37: /cgn2_6/ptodata/1/pna/US099D_COMB.seq:*
38: /cgn2_6/ptodata/1/pna/US100A_COMB.seq:*
39: /cgn2_6/ptodata/1/pna/US100B_COMB.seq:*
40: /cgn2_6/ptodata/1/pna/US101A_COMB.seq:*
41: /cgn2_6/ptodata/1/pna/US101B_COMB.seq:*
42: /cgn2_6/ptodata/1/pna/US102A_COMB.seq:*
43: /cgn2_6/ptodata/1/pna/US102B_COMB.seq:*
44: /cgn2_6/ptodata/1/pna/US6000_COMB.seq:*
45: /cgn2_6/ptodata/1/pna/US6001_COMB.seq:*
46: /cgn2_6/ptodata/1/pna/US6002_COMB.seq:*
47: /cgn2_6/ptodata/1/pna/US6003_COMB.seq:*
48: /cgn2_6/ptodata/1/pna/US6004_COMB.seq:*
49: /cgn2_6/ptodata/1/pna/US6005_COMB.seq:*
50: /cgn2_6/ptodata/1/pna/US6006_COMB.seq:*
51: /cgn2_6/ptodata/1/pna/US6007_COMB.seq:*
52: /cgn2_6/ptodata/1/pna/US6008_COMB.seq:*
53: /cgn2_6/ptodata/1/pna/US6009_COMB.seq:*
54: /cgn2_6/ptodata/1/pna/US6010_COMB.seq:*
55: /cgn2_6/ptodata/1/pna/US6011_COMB.seq:*
56: /cgn2_6/ptodata/1/pna/US6012_COMB.seq:*
57: /cgn2_6/ptodata/1/pna/US6013_COMB.seq:*
58: /cgn2_6/ptodata/1/pna/US6014_COMB.seq:*
59: /cgn2_6/ptodata/1/pna/US6015_COMB.seq:*
60: /cgn2_6/ptodata/1/pna/US6016_COMB.seq:*
61: /cgn2_6/ptodata/1/pna/US6017_COMB.seq:*
62: /cgn2_6/ptodata/1/pna/US6018_COMB.seq:*
63: /cgn2_6/ptodata/1/pna/US6019_COMB.seq:*
64: /cgn2_6/ptodata/1/pna/US6020_COMB.seq:*
65: /cgn2_6/ptodata/1/pna/US6021_COMB.seq:*
66: /cgn2_6/ptodata/1/pna/US6022_COMB.seq:*
67: /cgn2_6/ptodata/1/pna/US6023_COMB.seq:*
68: /cgn2_6/ptodata/1/pna/US6024_COMB.seq:*
69: /cgn2_6/ptodata/1/pna/US6025_COMB.seq:*
70: /cgn2_6/ptodata/1/pna/US6026_COMB.seq:*
71: /cgn2_6/ptodata/1/pna/US6027_COMB.seq:*
72: /cgn2_6/ptodata/1/pna/US6028_COMB.seq:*
73: /cgn2_6/ptodata/1/pna/US6029_COMB.seq:*
74: /cgn2_6/ptodata/1/pna/US6030_COMB.seq:*
75: /cgn2_6/ptodata/1/pna/US6031_COMB.seq:*
76: /cgn2_6/ptodata/1/pna/US6032_COMB.seq:*
77: /cgn2_6/ptodata/1/pna/US6033_COMB.seq:*
78: /cgn2_6/ptodata/1/pna/US6034_COMB.seq:*
79: /cgn2_6/ptodata/1/pna/US6035_COMB.seq:*
80: /cgn2_6/ptodata/1/pna/US6036_COMB.seq:*
81: /cgn2_6/ptodata/1/pna/US6037_COMB.seq:*
82: /cgn2_6/ptodata/1/pna/US6038_COMB.seq:*
83: /cgn2_6/ptodata/1/pna/US6039_COMB.seq:*
84: /cgn2_6/ptodata/1/pna/US6040_COMB.seq:*
85: /cgn2_6/ptodata/1/pna/US6041_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	402	100.0	402	25	US-09-654-617-457554	Sequence 457554,
	2	402	100.0	402	27	US-09-684-016-457554	Sequence 457554,
	3	402	100.0	402	32	US-09-850-147-1	Sequence 1, Appli
	4	402	100.0	402	64	US-60-202-213-1	Sequence 1, Appli
	5	266.8	66.4	426	32	US-09-850-147-40	Sequence 40, Appl
	6	266.8	66.4	426	64	US-60-202-213-40	Sequence 40, Appl
	7	266.8	66.4	456	25	US-09-654-617-451847	Sequence 451847,
	8	266.8	66.4	456	27	US-09-684-016-451847	Sequence 451847,
	9	264	65.7	433	32	US-09-850-147-1092	Sequence 1092, Ap
	10	264	65.7	433	64	US-60-202-213-1089	Sequence 1089, Ap
	11	255	63.4	385	32	US-09-850-147-10	Sequence 10, Appl
	12	255	63.4	385	64	US-60-202-213-10	Sequence 10, Appl
	13	223.6	55.6	1446	28	US-09-708-427-52180	Sequence 52180, A
	14	223.4	55.6	1444	19	US-09-513-996A-50583	Sequence 50583, A
	15	221.4	55.1	1694	25	US-09-654-617-262628	Sequence 262628,
	16	221.4	55.1	1694	27	US-09-684-016-262628	Sequence 262628,
c	17	220.6	54.9	570	33	US-09-865-439A-116358	Sequence 116358,
c	18	220	54.7	484	33	US-09-865-419A-15731	Sequence 15731, A
c	19	220	54.7	484	64	US-60-208-063-2535	Sequence 2535, Ap
c	20	220	54.7	545	33	US-09-873-402A-41873	Sequence 41873, A
c	21	220	54.7	550	33	US-09-865-419A-51912	Sequence 51912, A

c	22	220	54.7	559	33	US-09-873-402A-54657	Sequence 54657, A
c	23	220	54.7	559	64	US-60-209-830-46016	Sequence 46016, A
	24	220	54.7	995	42	US-10-219-999-1500	Sequence 1500, Ap
	25	220	54.7	1325	42	US-10-219-999-18314	Sequence 18314, A
c	26	217	54.0	472	33	US-09-873-402A-11200	Sequence 11200, A
c	27	217	54.0	472	64	US-60-209-830-11200	Sequence 11200, A
c	28	213.6	53.1	450	64	US-60-207-458-54563	Sequence 54563, A
c	29	213	53.0	535	33	US-09-865-439A-115458	Sequence 115458,
	30	213	53.0	1480	42	US-10-219-999-22880	Sequence 22880, A
	31	208.2	51.8	1231	42	US-10-219-999-20024	Sequence 20024, A
c	32	205.6	51.1	553	33	US-09-865-439A-80117	Sequence 80117, A
c	33	205.6	51.1	553	64	US-60-207-458-124259	Sequence 124259,
c	34	202.4	50.3	479	33	US-09-865-419A-42700	Sequence 42700, A
	35	201.6	50.1	393	22	US-09-565-306-82924	Sequence 82924, A
c	36	196.4	48.9	397	64	US-60-207-458-43952	Sequence 43952, A
c	37	195	48.5	388	33	US-09-865-439A-18142	Sequence 18142, A
c	38	193	48.0	501	64	US-60-207-458-25263	Sequence 25263, A
c	39	192.6	47.9	411	33	US-09-865-419A-9944	Sequence 9944, Ap
c	40	184.2	45.8	351	33	US-09-865-439A-4361	Sequence 4361, Ap
c	41	181.8	45.2	500	33	US-09-865-439A-93679	Sequence 93679, A
c	42	181.8	45.2	500	64	US-60-207-458-137694	Sequence 137694,
c	43	179	44.5	396	33	US-09-865-439A-89549	Sequence 89549, A
c	44	179	44.5	396	64	US-60-207-458-133564	Sequence 133564,
	45	173.4	43.1	520	30	US-09-795-301-466	Sequence 466, App

ALIGNMENTS

```
RESULT 1
US-09-654-617-457554
; Sequence 457554, Application US/09654617
; GENERAL INFORMATION:
;  APPLICANT:       Kovalic, David K.
;  APPLICANT:  Liu, Jingdong
;  TITLE OF INVENTION:       Annotated Plant Genes
;  FILE REFERENCE:       38-21(15097)D
;  CURRENT APPLICATION NUMBER: US/09/654,617
;  CURRENT FILING DATE:  2000-09-05
;  NUMBER OF SEQ ID NOS:       463173
; SEQ ID NO 457554
;   LENGTH: 402
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
US-09-654-617-457554

  Query Match             100.0%;  Score 402;  DB 25;  Length 402;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-109;
  Matches  402;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CCGGGCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CCGGGCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGG 60

Qy     61 CCTTGAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCTTGAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTAT 120

Qy    121 ATTAAGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ATTAAGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAA 180

Qy    181 ACAACCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGAC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ACAACCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGAC 240

Qy    241 AATTTTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AATTTTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATT 300

Qy    301 GTGGAGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GTGGAGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACA 360

Qy    361 AGTACAAAAAACCAACGGGAGGGGCATAACTTCCTGTGTTCC 402
          |||||||||||||||||||||||||||||||||||||||||||
Db    361 AGTACAAAAAACCAACGGGAGGGGCATAACTTCCTGTGTTCC 402


RESULT 2
US-09-684-016-457554
; Sequence 457554, Application US/09684016
; GENERAL INFORMATION:
;  APPLICANT:       Kovalic, David K.
;  APPLICANT:  Liu, Jingdong
;  TITLE OF INVENTION:       Annotated Plant Genes
;  FILE REFERENCE:       38-21(15097)D
;  CURRENT APPLICATION NUMBER: US/09/684,016
;  CURRENT FILING DATE:  2000-10-10
;  PRIOR APPLICATION NUMBER:       US 09/654,617
;  PRIOR FILING DATE:       2000-09-05
;  NUMBER OF SEQ ID NOS:       463173
; SEQ ID NO 457554
;   LENGTH: 402
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
US-09-684-016-457554

  Query Match             100.0%;  Score 402;  DB 27;  Length 402;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-109;
  Matches  402;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CCGGGCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CCGGGCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGG 60

Qy     61 CCTTGAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCTTGAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTAT 120

Qy    121 ATTAAGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ATTAAGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAA 180

Qy    181 ACAACCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGAC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ACAACCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGAC 240

Qy    241 AATTTTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AATTTTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATT 300

Qy    301 GTGGAGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GTGGAGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACA 360

Qy    361 AGTACAAAAAACCAACGGGAGGGGCATAACTTCCTGTGTTCC 402
          |||||||||||||||||||||||||||||||||||||||||||
Db    361 AGTACAAAAAACCAACGGGAGGGGCATAACTTCCTGTGTTCC 402


RESULT 3
US-09-850-147-1
; Sequence 1, Application US/09850147
; GENERAL INFORMATION:
;  APPLICANT: Andersen, Scott E.
;  APPLICANT:  Edgerton, Michael D
;  TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(51914)B
;  CURRENT APPLICATION NUMBER: US/09/850,147
;  CURRENT FILING DATE:  2001-05-08
;  PRIOR APPLICATION NUMBER: US 60/202,213
;  PRIOR FILING DATE: 2000-05-08
;  PRIOR APPLICATION NUMBER: US 09/654,617
;  PRIOR FILING DATE: 2000-09-05
```

```
;  PRIOR APPLICATION NUMBER: US 09/684,016
;  PRIOR FILING DATE: 2000-10-10
;  NUMBER OF SEQ ID NOS: 18014
; SEQ ID NO 1
;   LENGTH: 402
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   OTHER INFORMATION: Clone ID: LIB3476-001-Q6-K1-A3
US-09-850-147-1

  Query Match             100.0%;  Score 402;  DB 32;  Length 402;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-109;
  Matches  402;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CCGGGCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CCGGGCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGG 60

Qy     61 CCTTGAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCTTGAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTAT 120

Qy    121 ATTAAGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ATTAAGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAA 180

Qy    181 ACAACCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGAC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ACAACCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGAC 240

Qy    241 AATTTTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AATTTTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATT 300

Qy    301 GTGGAGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GTGGAGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACA 360

Qy    361 AGTACAAAAAACCAACGGGAGGGGCATAACTTCCTGTGTTCC 402
          |||||||||||||||||||||||||||||||||||||||||||
Db    361 AGTACAAAAAACCAACGGGAGGGGCATAACTTCCTGTGTTCC 402


RESULT 4
US-60-202-213-1
; Sequence 1, Application US/60202213
; GENERAL INFORMATION:
;  APPLICANT:       Andersen, Scott E.
;  APPLICANT:  Edgerton, Michael D
;  TITLE OF INVENTION:       NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
;  TITLE OF INVENTION:  PLANTS
;  FILE REFERENCE:       38-21(51914)A
;  CURRENT APPLICATION NUMBER: US/60/202,213
;  CURRENT FILING DATE:  2000-05-08
;  NUMBER OF SEQ ID NOS:       17986
; SEQ ID NO 1
;   LENGTH: 402
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   OTHER INFORMATION: Clone ID: LIB3476-001-Q6-K1-A3
US-60-202-213-1

  Query Match             100.0%;  Score 402;  DB 64;  Length 402;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-109;
  Matches  402;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CCGGGCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CCGGGCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGG 60

Qy     61 CCTTGAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCTTGAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTAT 120

Qy    121 ATTAAGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ATTAAGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAA 180

Qy    181 ACAACCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGAC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ACAACCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGAC 240

Qy    241 AATTTTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AATTTTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATT 300

Qy    301 GTGGAGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GTGGAGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACA 360

Qy    361 AGTACAAAAAACCAACGGGAGGGGCATAACTTCCTGTGTTCC 402
          |||||||||||||||||||||||||||||||||||||||||||
Db    361 AGTACAAAAAACCAACGGGAGGGGCATAACTTCCTGTGTTCC 402


RESULT 5
US-09-850-147-40
; Sequence 40, Application US/09850147
; GENERAL INFORMATION:
;  APPLICANT: Andersen, Scott E.
;  APPLICANT:  Edgerton, Michael D
;  TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(51914)B
;  CURRENT APPLICATION NUMBER: US/09/850,147
;  CURRENT FILING DATE:  2001-05-08
;  PRIOR APPLICATION NUMBER: US 60/202,213
;  PRIOR FILING DATE: 2000-05-08
;  PRIOR APPLICATION NUMBER: US 09/654,617
;  PRIOR FILING DATE: 2000-09-05
;  PRIOR APPLICATION NUMBER: US 09/684,016
;  PRIOR FILING DATE: 2000-10-10
;  NUMBER OF SEQ ID NOS: 18014
; SEQ ID NO 40
;   LENGTH: 426
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   OTHER INFORMATION: Clone ID: LIB3476-004-Q6-K1-A3
US-09-850-147-40

  Query Match             66.4%;  Score 266.8;  DB 32;  Length 426;
  Best Local Similarity   83.8%;  Pred. No. 5.3e-69;
  Matches  337;  Conservative    0;  Mismatches   62;  Indels    3;  Gaps    3;

Qy      1 CCGGGCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGG 60
          || || |||||||||||||||| ||||  | |  | ||  |||||| ||||||||| |||
Db     25 CCCGGGCGGGCTCACCAGCTGGTCCAGGAAGGCTTAAAGTACATGTTCCAGGAGAA-GGG 83

Qy     61 CCTTGAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTAT 120
          |||||  |  ||||||||||||||  | |  |     |||||||||||||| |||||||||
Db     84 CCTTGTAACTGTGTGGTCTGCACCTAATTATTGCTACAGATGTGGCAATGTAGCTTCTAT 143

Qy    121 ATTAAGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAA 180
          |||||||||||||||||| ||||| |||||||||||| |||||||||||||| ||||| |
Db    144 ATTAAGCTTCGATGAGAAGATGGAAA-GAGATGTCAAGTTCTTCACAGAGACAGAGGAGA 202

Qy    181 ACAACCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGAC 240
          |||||||||| | ||| |||||| | || | |||||||||||||| ||||||||| || ||
Db    203 ACAACCAGATGCGAGGCCCAAGGAC-TGCAGTCCCATATTTCCTCTGACTTCAGTTGTAC 261

Qy    241 AATTTTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATT 300
          |||||||||| |||||||||||||| ||||| |||||| |||||||||   ||| |
Db    262 AATTTTAGAAGTCAATATTTAAATCTGTAAGGGTTAAAGATCAATCAGATGTGTTTATTG 321

Qy    301 GTGGAGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACA 360
           || |||||||||||||| ||| |||||||||||||| ||||||| |||||||||| ||||
```

```
Db    322 TTGTAGGTTCAGATCATAGTGCAATCTGATGCATTCGATTTCTGTTTTCTTTTGACCACA 381

Qy    361 AGTACAAAAAACCAACGGGAGGGGCATAACTTCCTGTGTTCC 402
          |||||| | |||| |||||| | |||| |||||||| ||||
Db    382 TGTACAATAGACCATCGGGAGTGTCATATCTTCCTGTATTCC 423


RESULT 6
US-60-202-213-40
; Sequence 40, Application US/60202213
; GENERAL INFORMATION:
;  APPLICANT:       Andersen, Scott E.
;  APPLICANT:  Edgerton, Michael D
;  TITLE OF INVENTION:       NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
;  TITLE OF INVENTION:  PLANTS
;  FILE REFERENCE:       38-21(51914)A
;  CURRENT APPLICATION NUMBER: US/60/202,213
;  CURRENT FILING DATE:  2000-05-08
;  NUMBER OF SEQ ID NOS:       17986
; SEQ ID NO 40
;   LENGTH: 426
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   OTHER INFORMATION: Clone ID: LIB3476-004-Q6-K1-A3
US-60-202-213-40

  Query Match             66.4%;  Score 266.8;  DB 64;  Length 426;
  Best Local Similarity   83.8%;  Pred. No. 5.3e-69;
  Matches  337;  Conservative    0;  Mismatches   62;  Indels    3;  Gaps    3;

Qy      1 CCGGGCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGG 60
          || || |||||||||||||||| |||| | | | ||  |||||| ||||||||| |||
Db     25 CCCGGGCGGGCTCACCAGCTGGTCCAGGAAGGCTTAAAGTACATGTTCCAGGAGAA-GGG 83

Qy     61 CCTTGAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTAT 120
          ||||| |  |||||||||||||||  | |  |    |||||||||||||| |||||||||
Db     84 CCTTGTAACTGTGTGGTCTGCACCTAATTATTGCTACAGATGTGGCAATGTAGCTTCTAT 143

Qy    121 ATTAAGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAA 180
          |||||||||||||||||| ||||| |||||||||| |||||||||||| ||||| |
Db    144 ATTAAGCTTCGATGAGAAGATGGAAA-GAGATGTCAAGTTCTTCACAGAGACAGAGGAGA 202

Qy    181 ACAACCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGAC 240
          |||||||||| | ||| |||||| | || | |||||||||||||| ||||||||  || ||
Db    203 ACAACCAGATGCGAGGCCCAAGGAC-TGCAGTCCCATATTTCCTCTGACTTCAGTTGTAC 261

Qy    241 AATTTTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATT 300
          |||||||||| |||||||||||||| ||||| |||||| ||||||||||   ||| |
Db    262 AATTTTAGAAGTCAATATTTAAATCTGTAAGGGTTAAAGATCAATCAGATGTGTTTATTG 321

Qy    301 GTGGAGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACA 360
          || |||||||||||||| ||| |||||||||||| ||||||| |||||||||| ||||
Db    322 TTGTAGGTTCAGATCATAGTGCAATCTGATGCATTCGATTTCTGTTTTCTTTTGACCACA 381

Qy    361 AGTACAAAAAACCAACGGGAGGGGCATAACTTCCTGTGTTCC 402
          |||||| | |||| |||||| | |||| |||||||| ||||
Db    382 TGTACAATAGACCATCGGGAGTGTCATATCTTCCTGTATTCC 423


RESULT 7
US-09-654-617-451847
; Sequence 451847, Application US/09654617
; GENERAL INFORMATION:
;  APPLICANT:       Kovalic, David K.
;  APPLICANT:  Liu, Jingdong
;  TITLE OF INVENTION:       Annotated Plant Genes
;  FILE REFERENCE:       38-21(15097)D
;  CURRENT APPLICATION NUMBER: US/09/654,617
;  CURRENT FILING DATE:  2000-09-05
;  NUMBER OF SEQ ID NOS:       463173
; SEQ ID NO 451847
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   OTHER INFORMATION: unsure at all n locations
US-09-654-617-451847

  Query Match             66.4%;  Score 266.8;  DB 25;  Length 456;
  Best Local Similarity   83.8%;  Pred. No. 5.4e-69;
  Matches  337;  Conservative    0;  Mismatches   62;  Indels    3;  Gaps    3;

Qy      1 CCGGGCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGG 60
          || || |||||||||||||||| |||| | | | ||  |||||| ||||||||| |||
Db     25 CCCGGGCGGGCTCACCAGCTGGTCCAGGAAGGCTTAAAGTACATGTTCCAGGAGAA-GGG 83

Qy     61 CCTTGAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTAT 120
          |||||  |||||||||||||||  | |  |    |||||||||||||| |||||||||
Db     84 CCTTGTAACTGTGTGGTCTGCACCTAATTATTGCTACAGATGTGGCAATGTAGCTTCTAT 143

Qy    121 ATTAAGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAA 180
          |||||||||| | ||| |||||| | || | |||||||||||||| ||||||||  || ||
Db    203 ACAACCAGATGCGAGGCCCAAGGAC-TGCAGTCCCATATTTCCTCTGACTTCAGTTGTAC 261

Qy    241 AATTTTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATT 300
          |||||||||| |||||||||||||| ||||| |||||| ||||||||||   ||| |
Db    262 AATTTTAGAAGTCAATATTTAAATCTGTAAGGGTTAAAGATCAATCAGATGTGTTTATTG 321

Qy    301 GTGGAGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACA 360
          || |||||||||||||| ||| |||||||||||| ||||||| |||||||||| ||||
Db    322 TTGTAGGTTCAGATCATAGTGCAATCTGATGCATTCGATTTCTGTTTTCTTTTGACCACA 381

Qy    361 AGTACAAAAAACCAACGGGAGGGGCATAACTTCCTGTGTTCC 402
          |||||| | |||| |||||| | |||| |||||||| ||||
Db    382 TGTACAATAGACCATCGGGAGTGTCATATCTTCCTGTATTCC 423


RESULT 8
US-09-684-016-451847
; Sequence 451847, Application US/09684016
; GENERAL INFORMATION:
;  APPLICANT:       Kovalic, David K.
;  APPLICANT:  Liu, Jingdong
;  TITLE OF INVENTION:       Annotated Plant Genes
;  FILE REFERENCE:       38-21(15097)D
;  CURRENT APPLICATION NUMBER: US/09/684,016
;  CURRENT FILING DATE:  2000-10-10
;  PRIOR APPLICATION NUMBER:       US 09/654,617
;  PRIOR FILING DATE:       2000-09-05
;  NUMBER OF SEQ ID NOS:       463173
; SEQ ID NO 451847
;   LENGTH: 456
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(456)
;   OTHER INFORMATION: unsure at all n locations
US-09-684-016-451847

  Query Match             66.4%;  Score 266.8;  DB 27;  Length 456;
  Best Local Similarity   83.8%;  Pred. No. 5.4e-69;
  Matches  337;  Conservative    0;  Mismatches   62;  Indels    3;  Gaps    3;

Qy      1 CCGGGCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGG 60
          || || |||||||||||||||| |||| | | | ||  |||||| ||||||||| |||
Db     25 CCCGGGCGGGCTCACCAGCTGGTCCAGGAAGGCTTAAAGTACATGTTCCAGGAGAA-GGG 83

Qy     61 CCTTGAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTAT 120
          |||||  |||||||||||||||  | |  |    |||||||||||||| |||||||||
Db     84 CCTTGTAACTGTGTGGTCTGCACCTAATTATTGCTACAGATGTGGCAATGTAGCTTCTAT 143

Qy    121 ATTAAGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAA 180
```

```
          |||||||||||||||||||| ||||| |||||||||| |||||||||||| ||||| |
Db    144 ATTAAGCTTCGATGAGAAGATGGAAA-GAGATGTCAAGTTCTTCACAGAGACAGAGGAGA 202

Qy    181 ACAACCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGAC 240
          |||||||||| | ||| |||||| | || | |||||||||||| ||||||||| || ||
Db    203 ACAACCAGATGCGAGGCCCAAGGAC-TGCAGTCCCATATTTCCTCTGACTTCAGTTGTAC 261

Qy    241 AATTTTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATT 300
          |||||||||||  |||||||||||||||  ||||| |||||| ||||||||||   ||| |
Db    262 AATTTTAGAAGTCAATATTTAAATCTGTAAGGGTTAAAGATCAATCAGATGTGTTTATTG 321

Qy    301 GTGGAGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACA 360
           || ||||||||||||||| ||| |||||||||||||| ||||||| |||||||||| ||||
Db    322 TTGTAGGTTCAGATCATAGTGCAATCTGATGCATTCGATTTCTGTTTTCTTTTGACCACA 381

Qy    361 AGTACAAAAAACCAACGGGAGGGGCATAACTTCCTGTGTTCC 402
           |||||| | |||| |||||| | |||| |||||||| ||||
Db    382 TGTACAATAGACCATCGGGAGTGTCATATCTTCCTGTATTCC 423


RESULT 9
US-09-850-147-1092
; Sequence 1092, Application US/09850147
; GENERAL INFORMATION:
;  APPLICANT: Andersen, Scott E.
;  APPLICANT:  Edgerton, Michael D
;  TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(51914)B
;  CURRENT APPLICATION NUMBER: US/09/850,147
;  CURRENT FILING DATE:  2001-05-08
;  PRIOR APPLICATION NUMBER: US 60/202,213
;  PRIOR FILING DATE: 2000-05-08
;  PRIOR APPLICATION NUMBER: US 09/654,617
;  PRIOR FILING DATE: 2000-09-05
;  PRIOR APPLICATION NUMBER: US 09/684,016
;  PRIOR FILING DATE: 2000-10-10
;  NUMBER OF SEQ ID NOS: 18014
; SEQ ID NO 1092
;   LENGTH: 433
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(433)
;   OTHER INFORMATION: unsure at all n locations
;   OTHER INFORMATION: Clone ID: LIB3476-004-Q6-K2-A3
US-09-850-147-1092

  Query Match             65.7%;  Score 264;  DB 32;  Length 433;
  Best Local Similarity   84.1%;  Pred. No. 3.7e-68;
  Matches  333;  Conservative    0;  Mismatches   60;  Indels    3;  Gaps    3;

Qy      7 CGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTTGA 66
          |||||||||||||||| ||||  | |  | ||  |||||| ||||||||| ||||||||
Db      1 CGGGCTCACCAGCTGGTCCAGGAAGGCTTAAAGTACATGTTCCAGGAGAA-GGGCCTTGT 59

Qy     67 ACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTAAG 126
          |  ||||||||||||||  | |  |    |||||||||||||| ||||||||||||||
Db     60 AACTGTGTGGTCTGCACCTAATTATTGCTACAGATGTGGCAATGTAGCTTCTATATTAAG 119

Qy    127 CTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAACC 186
          |||||||||||||| ||||| |||||||||| |||||||||||| ||||| |||||||
Db    120 CTTCGATGAGAAGATGGAAA-GAGATGTCAAGTTCTTCACAGAGACAGAGGAGAACAACC 178

Qy    187 AGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGACAATTTT 246
          |||| | ||| |||||| | || | |||||||||||| ||||||||| || ||||||||
Db    179 AGATGCGAGGCCCAAGGAC-TGCAGTCCCATATTTCCTCTGACTTCAGTTGTACAATTTT 237

Qy    247 AGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGGAG 306
          |||||  |||||||||||||| ||||| |||||| ||||||||||   ||| |  || ||
Db    238 AGAAGTCAATATTTAAATCTGTAAGGGTTAAAGATCAATCAGATGTGTTTATTGTTGTAG 297

Qy    307 GTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACAAGTACA 366
          |||||||||||  ||| |||||||||||||| ||||||| ||||||||| |||| |||||
Db    298 GTTCAGATCATAGTGCAATCTGATGCATTCGATTTCTGTTTTCTTTTGACCACATGTACA 357

Qy    367 AAAAACCAACGGGAGGGGCATAACTTCCTGTGTTCC 402
          | | |||| |||||| | |||| |||||||| ||||
Db    358 ATAGACCATCGGGAGTGTCATATCTTCCTGTATTCC 393


RESULT 10
US-60-202-213-1089
; Sequence 1089, Application US/60202213
; GENERAL INFORMATION:
;  APPLICANT:       Andersen, Scott E.
;  APPLICANT:  Edgerton, Michael D
;  TITLE OF INVENTION:       NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WI
;  TITLE OF INVENTION:  PLANTS
;  FILE REFERENCE:       38-21(51914)A
;  CURRENT APPLICATION NUMBER: US/60/202,213
;  CURRENT FILING DATE:  2000-05-08
;  NUMBER OF SEQ ID NOS:       17986
; SEQ ID NO 1089
;   LENGTH: 433
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   OTHER INFORMATION: Clone ID: LIB3476-004-Q6-K2-A3
US-60-202-213-1089

  Query Match             65.7%;  Score 264;  DB 64;  Length 433;
  Best Local Similarity   84.1%;  Pred. No. 3.7e-68;
  Matches  333;  Conservative    0;  Mismatches   60;  Indels    3;  Gaps    3;

Qy      7 CGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTTGA 66
          |||||||||||||||| ||||  | |  | ||  |||||| ||||||||| ||||||||
Db      1 CGGGCTCACCAGCTGGTCCAGGAAGGCTTAAAGTACATGTTCCAGGAGAA-GGGCCTTGT 59

Qy     67 ACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTAAG 126
          |  ||||||||||||||  | |  |    |||||||||||||| ||||||||||||||
Db     60 AACTGTGTGGTCTGCACCTAATTATTGCTACAGATGTGGCAATGTAGCTTCTATATTAAG 119

Qy    127 CTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAACC 186
          |||||||||||||| ||||| |||||||||| |||||||||||| ||||| |||||||
Db    120 CTTCGATGAGAAGATGGAAA-GAGATGTCAAGTTCTTCACAGAGACAGAGGAGAACAACC 178

Qy    187 AGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGACAATTTT 246
          |||| | ||| |||||| | || | |||||||||||| ||||||||| || ||||||||
Db    179 AGATGCGAGGCCCAAGGAC-TGCAGTCCCATATTTCCTCTGACTTCAGTTGTACAATTTT 237

Qy    247 AGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGGAG 306
          |||||  |||||||||||||| ||||| |||||| ||||||||||   ||| |  || ||
Db    238 AGAAGTCAATATTTAAATCTGTAAGGGTTAAAGATCAATCAGATGTGTTTATTGTTGTAG 297

Qy    307 GTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACAAGTACA 366
          |||||||||||  ||| |||||||||||||| ||||||| ||||||||| |||| |||||
Db    298 GTTCAGATCATAGTGCAATCTGATGCATTCGATTTCTGTTTTCTTTTGACCACATGTACA 357

Qy    367 AAAAACCAACGGGAGGGGCATAACTTCCTGTGTTCC 402
          | | |||| |||||| | |||| |||||||| ||||
Db    358 ATAGACCATCGGGAGTGTCATATCTTCCTGTATTCC 393


RESULT 11
US-09-850-147-10
; Sequence 10, Application US/09850147
; GENERAL INFORMATION:
;  APPLICANT: Andersen, Scott E.
;  APPLICANT:  Edgerton, Michael D
;  TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(51914)B
;  CURRENT APPLICATION NUMBER: US/09/850,147
;  CURRENT FILING DATE:  2001-05-08
;  PRIOR APPLICATION NUMBER: US 60/202,213
```

```
;  PRIOR FILING DATE: 2000-05-08
;  PRIOR APPLICATION NUMBER: US 09/654,617
;  PRIOR FILING DATE: 2000-09-05
;  PRIOR APPLICATION NUMBER: US 09/684,016
;  PRIOR FILING DATE: 2000-10-10
;  NUMBER OF SEQ ID NOS: 18014
; SEQ ID NO 10
;   LENGTH: 385
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   OTHER INFORMATION: Clone ID: LIB3476-001-Q6-K2-A3
US-09-850-147-10

  Query Match             63.4%;  Score 255;  DB 32;  Length 385;
  Best Local Similarity   83.7%;  Pred. No. 1.8e-65;
  Matches  324;  Conservative    0;  Mismatches   60;  Indels    3;  Gaps    3;

Qy      16 CAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTTGAACTTGTGTG 75
           ||||||| ||||  | |  | ||  ||||||  |||||||||||| | | ||||| | |||
Db       1 CAGCTGGGCCAGGAAGGCTTAAAGTACATGTTCCAGGAGAAGGGCCTTGGAACTGG-GTG 59

Qy      76 GTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTAAGCTTCGATGA 135
           |||||||||  | |  |    |||||||||||||| |||||||||||||||||||||||
Db      60 GTCTGCACCTAATTATTGCTACAGATGTGGCAATGTAGCTTCTATATTAAGCTTCGATGA 119

Qy     136 GAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAACCAGATTCCAG 195
           ||||| ||||| |||||||||||| |||||||||||||||||||| |||||||||||  | ||
Db     120 GAAGATGGAAA-GAGATGTCAAGTTCTTCACAGAGACCGAGGAGAACAACCAGATGCGAG 178

Qy     196 GGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGACAATTTTAGAAGGCAA 255
           | |||||| | || | ||||||||||||| |||||||||  || ||||||||||||| |||
Db     179 GCCCAAGGAC-TGCAGTCCCATATTTCCTCTGACTTCAGTTGTACAATTTTAGAAGTCAA 237

Qy     256 TATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGGAGGTTCAGATC 315
           |||||||||||||||||| |||||| ||||||||||   ||| |  || ||||||||||
Db     238 TATTTAAATCTGGAAGGGTTAAAGATCAATCAGATGTGTTTATTGTTGTAGGTTCAGATC 297

Qy     316 ATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACAAGTACAAAAAACCAA 375
           ||| ||| | ||||||||||| ||||||| |||||||||| |||| |||||| | ||||
Db     298 ATAGTGCAAGCTGATGCATTCGATTTCTGTTTTCTTTTGACCACATGTACAATAGACCAT 357

Qy     376 CGGGAGGGGCATAACTTCCTGTGTTCC 402
           ||||||| | |||| |||||||| ||||
Db     358 CGGGAGTGTCATATCTTCCTGTATTCC 384


RESULT 12
US-60-202-213-10
; Sequence 10, Application US/60202213
; GENERAL INFORMATION:
;  APPLICANT:       Andersen, Scott E.
;  APPLICANT:  Edgerton, Michael D
;  TITLE OF INVENTION:       NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
;  TITLE OF INVENTION:  PLANTS
;  FILE REFERENCE:       38-21(51914)A
;  CURRENT APPLICATION NUMBER: US/60/202,213
;  CURRENT FILING DATE:  2000-05-08
;  NUMBER OF SEQ ID NOS:       17986
; SEQ ID NO 10
;   LENGTH: 385
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   OTHER INFORMATION: Clone ID: LIB3476-001-Q6-K2-A3
US-60-202-213-10

  Query Match             63.4%;  Score 255;  DB 64;  Length 385;
  Best Local Similarity   83.7%;  Pred. No. 1.8e-65;
  Matches  324;  Conservative    0;  Mismatches   60;  Indels    3;  Gaps    3;

Qy      16 CAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTTGAACTTGTGTG 75
           |||||||  ||||  | |  | ||  ||||||  |||||||||||| | | ||||| | |||
Db       1 CAGCTGGGCCAGGAAGGCTTAAAGTACATGTTCCAGGAGAAGGGCCTTGGAACTGG-GTG 59

Qy      76 GTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTAAGCTTCGATGA 135
           |||||||||  | |  |    |||||||||||||| |||||||||||||||||||||||
Db      60 GTCTGCACCTAATTATTGCTACAGATGTGGCAATGTAGCTTCTATATTAAGCTTCGATGA 119

Qy     136 GAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAACCAGATTCCAG 195
           ||||| ||||| |||||||||||| |||||||||||||||||||| |||||||||||  | ||
Db     120 GAAGATGGAAA-GAGATGTCAAGTTCTTCACAGAGACCGAGGAGAACAACCAGATGCGAG 178

Qy     196 GGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGACAATTTTAGAAGGCAA 255
           | |||||| | || | ||||||||||||| |||||||||  || ||||||||||||| |||
Db     179 GCCCAAGGAC-TGCAGTCCCATATTTCCTCTGACTTCAGTTGTACAATTTTAGAAGTCAA 237

Qy     256 TATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGGAGGTTCAGATC 315
           |||||||||||||||||| |||||| ||||||||||   ||| |  || ||||||||||
Db     238 TATTTAAATCTGGAAGGGTTAAAGATCAATCAGATGTGTTTATTGTTGTAGGTTCAGATC 297

Qy     316 ATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACAAGTACAAAAAACCAA 375
           ||| ||| | ||||||||||| ||||||| |||||||||| |||| |||||| | ||||
Db     298 ATAGTGCAAGCTGATGCATTCGATTTCTGTTTTCTTTTGACCACATGTACAATAGACCAT 357

Qy     376 CGGGAGGGGCATAACTTCCTGTGTTCC 402
           ||||||| | |||| |||||||| ||||
Db     358 CGGGAGTGTCATATCTTCCTGTATTCC 384
```

```
Qy      76 GTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTAAGCTTCGATGA 135
           |||||||||  | |  |    |||||||||||||| |||||||||||||||||||||||
Db      60 GTCTGCACCTAATTATTGCTACAGATGTGGCAATGTAGCTTCTATATTAAGCTTCGATGA 119

Qy     136 GAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAACCAGATTCCAG 195
           ||||| ||||| |||||||||||| |||||||||||||||||||| |||||||||||  | ||
Db     120 GAAGATGGAAA-GAGATGTCAAGTTCTTCACAGAGACCGAGGAGAACAACCAGATGCGAG 178

Qy     196 GGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGACAATTTTAGAAGGCAA 255
           | |||||| | || | ||||||||||||| |||||||||  || ||||||||||||| |||
Db     179 GCCCAAGGAC-TGCAGTCCCATATTTCCTCTGACTTCAGTTGTACAATTTTAGAAGTCAA 237

Qy     256 TATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGGAGGTTCAGATC 315
           |||||||||||||||||| |||||| ||||||||||   ||| |  || ||||||||||
Db     238 TATTTAAATCTGGAAGGGTTAAAGATCAATCAGATGTGTTTATTGTTGTAGGTTCAGATC 297

Qy     316 ATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACAAGTACAAAAAACCAA 375
           ||| ||| | ||||||||||| ||||||| |||||||||| |||| |||||| | ||||
Db     298 ATAGTGCAAGCTGATGCATTCGATTTCTGTTTTCTTTTGACCACATGTACAATAGACCAT 357

Qy     376 CGGGAGGGGCATAACTTCCTGTGTTCC 402
           ||||||| | |||| |||||||| ||||
Db     358 CGGGAGTGTCATATCTTCCTGTATTCC 384


RESULT 13
US-09-708-427-52180
; Sequence 52180, Application US/09708427
; GENERAL INFORMATION:
;  APPLICANT: N. ALEXANDROV et al.
;  TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
;  TITLE OF INVENTION:  THEREBY
;  FILE REFERENCE: 2750-1243P
;  CURRENT APPLICATION NUMBER: US/09/708,427
;  CURRENT FILING DATE:  2000-11-09
;  NUMBER OF SEQ ID NOS: 85364
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52180
;   LENGTH: 1446
;   TYPE: DNA
;   ORGANISM: Zea mays subsp. mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1..1446
;   OTHER INFORMATION: any n = a, g, c, t, unknown, or other
;   NAME/KEY: misc_feature
;   LOCATION: 1..1446
;   OTHER INFORMATION: Ceres Seq. ID 1928446
US-09-708-427-52180

  Query Match             55.6%;  Score 223.6;  DB 28;  Length 1446;
  Best Local Similarity   77.0%;  Pred. No. 8.2e-56;
  Matches  305;  Conservative    4;  Mismatches   83;  Indels    4;  Gaps    3;

Qy       4 GGCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCT 63
           | ::|||||||||||||||| ||||  | |  | |:  ||||||  |  |   |:||||||
Db    1005 GCSSGGGCTCACCAGCTGGTCCAGGAAGGCTTAARGTACATGTTGCCAGGAGARGGGCCT 1064

Qy      64 TGAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATT 123
           ||    |||||||||||-||||  | |  |    |||||||||||||  |||||||| ||
Db    1065 TGTCACTGTGTGGTCTGCACCTAATTATTGCTACAGATGTGGCAATGTCGCTTCTATCTT 1124

Qy     124 AAGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACA 183
           ||||||||||||||||| |||||||  |||||||| ||||||||||||  ||||| ||||
Db    1125 AAGCTTCGATGAGAAGATGGAAAGG-GATGTCAAGTTCTTCACAGAGACAGAGGAGAACA 1183

Qy     184 ACCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGACAAT 243
           ||||||| | ||| |||||| |  | | ||||||||||||| ||||||||  || |||||
Db    1184 ACCAGATGCGAGGCCCAAGGAC-CGCAGTCCCATATTTCCTATGACTTCAATTGTACAAT 1242

Qy     244 TTTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTG 303
           |||||||| |  |||| |||||| ||||| |||||| ||||||||||   ||| |  |
Db    1243 TTTAGAAGTC--TATTCGAATCTGTAAGGGCTAAAGATCAATCAGATGTGTTTATTGTTT 1300
```

```
Qy    304 GAGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACAAGT 363
          |||| ||||| ||| ||| ||||||||| ||  ||||||| |||| ||| | |||| ||
Db   1301 TAGGTCCAGATTATAGTGCAATCTGATGCGTTTGATTTCTGTTTTCCTTTAACCACATGT 1360

Qy    364 ACAAAAAACCAACGGGAGGGGCATAACTTCCTGTGT 399
          |||| | |||| |||||  | ||||  ||||||| |
Db   1361 ACAATAGACCATCGGGACAGTCATATGTTCCTGTAT 1396


RESULT 14
US-09-513-996A-50583
; Sequence 50583, Application US/09513996A
; GENERAL INFORMATION:
;  APPLICANT: N. ALEXANDROV et al.
;  TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
;  TITLE OF INVENTION:  ENCODED THEREBY
;  FILE REFERENCE: 2750-709P
;  CURRENT APPLICATION NUMBER: US/09/513,996A
;  CURRENT FILING DATE:  2000-02-25
;  NUMBER OF SEQ ID NOS: 81028
; SEQ ID NO 50583
;   LENGTH: 1444
;   TYPE: DNA
;   ORGANISM: Zea mays subsp. mays
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: 1..1444
;   OTHER INFORMATION: any n or Xaa = unknown
;   FEATURE:
;   OTHER INFORMATION: Location 1..1444 / Ceres Seq. ID 2065831
US-09-513-996A-50583

  Query Match             55.6%;  Score 223.4;  DB 19;  Length 1444;
  Best Local Similarity   79.5%;  Pred. No. 9.4e-56;
  Matches  314;  Conservative    0;  Mismatches   76;  Indels    5;  Gaps    4;

Qy      5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
          |||||||||||||||||| ||||  | |  | ||  |||||| ||||||||| |||||||
Db   1005 GCCGGGCTCACCAGCTGGTCCAGGAAGGCTTAAAGTACATGTTCCAGGAGAA-GGGCCTT 1063

Qy     65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
          |    ||||||||||||||||  | |  |    ||||||||||||||  |||||||| |||
Db   1064 GTCACTGTGTGGTCTGCACCTAATTATTGCTACAGATGTGGCAATGTCGCTTCTATCTTA 1123

Qy    125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
          |||||||||||||||| ||||||| ||||||||| |||||||||||||| ||||| |||||
Db   1124 AGCTTCGATGAGAAGATGGAAAGG-GATGTCAAGTTCTTCACAGAGACAGAGGAGAACAA 1182

Qy    185 CCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGACAATT 244
          ||||||  | ||| ||||||  |  | | |||||||||||||| ||||||||  || ||||||
Db   1183 CCAGATGCGAGGCCCAAGGAC-CGCAGTCCCATATTTCCTATGACTTCAATTGTACAATT 1241

Qy    245 TTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGG 304
          |||||||  |  ||||  |||||| ||||| |||||| ||||||||||   ||| |  |
Db   1242 TTAGAAGTC--TATTCGAATCTGTAAGGGCTAAAGATCAATCAGATGTGTTTATTGTTTT 1299

Qy    305 AGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACAAGTA 364
          |||| ||||| ||| ||| |||||||||| ||  ||||||| |||| ||| | |||| |||
Db   1300 AGGTCCAGATTATAGTGCAATCTGATGCGTTTGATTTCTGTTTTCCTTTAACCACATGTA 1359

Qy    365 CAAAAAACCAACGGGAGGGGCATAACTTCCTGTGT 399
          ||| | |||| |||||  | ||||  ||||||| |
Db   1360 CAATAGACCATCGGGACAGTCATATGTTCCTGTAT 1394


RESULT 15
US-09-654-617-262628
; Sequence 262628, Application US/09654617
; GENERAL INFORMATION:
;  APPLICANT:       Kovalic, David K.
;  APPLICANT:  Liu, Jingdong
;  TITLE OF INVENTION:       Annotated Plant Genes
;  FILE REFERENCE:       38-21(15097)D
;  CURRENT APPLICATION NUMBER: US/09/654,617
;  CURRENT FILING DATE:  2000-09-05
;  NUMBER OF SEQ ID NOS:       463173
; SEQ ID NO 262628
;   LENGTH: 1694
;   TYPE: DNA
;   ORGANISM: Zea mays
US-09-654-617-262628

  Query Match             55.1%;  Score 221.4;  DB 25;  Length 1694;
  Best Local Similarity   78.7%;  Pred. No. 4e-55;
  Matches  314;  Conservative    0;  Mismatches   81;  Indels    4;  Gaps    4;

Qy      5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
          |||||||||||||||||| ||||  | |  | ||  |||||| ||||||||| |||||||
Db    979 GCCGGGCTCACCAGCTGGTCCAGGAAGGCTTAAAGTACATGTTCCAGGAGAA-GGGCCTT 1037

Qy     65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
          |    ||||||||||||||  | |  |    ||||||||||||||  |||||||||||
Db   1038 GTCACTGTGTGGTCTGCACCTAATTATTGCTACAGATGTGGCAATGTCGCTTCTATATTA 1097

Qy    125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
          |||||||||||||||| ||||| |||||||||| |||||||||||| ||||| |||||
Db   1098 AGCTTCGATGAGAAGATGGAAA-GAGATGTCAAGTTCTTCACAGAGACAGAGGAGAACAA 1156

Qy    185 CCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGACAATT 244
          ||||||  | ||| ||||||  |  | | |||||||||||||| ||||||||  || ||||||
Db   1157 CCAGATGCGAGGCCCAAGGAC-CGCAGTCCCATATTTCCTCTGACTTCAATTGTACAATT 1215

Qy    245 TTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGG 304
          |||||||  | |  ||| ||||||  ||||| ||||||  || |||||||   ||| |  ||
Db   1216 TTAGAAGTCTACATTCGAATCTGTAAGGGCTAAAGATCAGTCAGATGTGTTTATTGTTGT 1275

Qy    305 AGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTT-CTTTTGAACACAAGT 363
          |||| |||||||| ||| |||||||| ||  ||||||  ||| ||||| | |||| ||
Db   1276 AGGTCCAGATCATAGTGCAATCTGATGCGTTTGATTTCTCTTTTCCTTTTAACCACATGT 1335

Qy    364 ACAAAAAACCAACGGGAGGGGCATAACTTCCTGTGTTCC 402
          |||| | |||| | |||  | |  |   ||||||  | ||
Db   1336 ACAATAGACCATCCGGACAGTCTCATGCTCCTGTATCCC 1374


Search completed: January  2, 2003, 14:25:02
Job time : 3766 secs
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: January 2, 2003, 12:39:27 ; Search time 146 Seconds
(without alignments)
2287.313 Million cell updates/sec

Title: US-09-850-147-1
Perfect score: 402
Sequence: 1 ccgggccgggctcaccagct.....ggcataacttcctgtgttcc 402

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 655294 seqs, 415357811 residues

Total number of hits satisfying chosen parameters: 1310588

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	88.8	22.1	533	5	US-09-531-113-41876	Sequence 41876, A
	2	88	21.9	1184	1	PCT-US02-28445-13	Sequence 13, Appl
	3	88	21.9	1184	6	US-10-236-699-13	Sequence 13, Appl
c	4	85.6	21.3	539	5	US-09-531-113-2951	Sequence 2951, Ap
	5	61.2	15.2	1014	1	PCT-US02-28445-5	Sequence 5, Appli
	6	61.2	15.2	1014	6	US-10-236-699-5	Sequence 5, Appli
c	7	33.8	8.4	183	6	US-10-092-411A-981	Sequence 981, App
c	8	33.8	8.4	1291	5	US-09-724-676-16911	Sequence 16911, A
c	9	33.8	8.4	1291	5	US-09-724-676A-16911	Sequence 16911, A
c	10	33.8	8.4	1317	5	US-09-724-676-16914	Sequence 16914, A
c	11	33.8	8.4	1317	5	US-09-724-676A-16914	Sequence 16914, A
c	12	33.8	8.4	1358	5	US-09-724-676-16912	Sequence 16912, A
c	13	33.8	8.4	1358	5	US-09-724-676A-16912	Sequence 16912, A
c	14	33.8	8.4	1384	5	US-09-724-676-16915	Sequence 16915, A
c	15	33.8	8.4	1384	5	US-09-724-676A-16915	Sequence 16915, A
c	16	33.8	8.4	1388	5	US-09-724-676-16910	Sequence 16910, A
c	17	33.8	8.4	1388	5	US-09-724-676A-16910	Sequence 16910, A
c	18	33.8	8.4	1414	5	US-09-724-676-16913	Sequence 16913, A
c	19	33.8	8.4	1414	5	US-09-724-676A-16913	Sequence 16913, A
c	20	33.8	8.4	1544	5	US-09-724-676-16905	Sequence 16905, A
c	21	33.8	8.4	1544	5	US-09-724-676A-16905	Sequence 16905, A
c	22	33.8	8.4	1570	5	US-09-724-676-16908	Sequence 16908, A
c	23	33.8	8.4	1570	5	US-09-724-676A-16908	Sequence 16908, A
c	24	33.8	8.4	1611	5	US-09-724-676-16906	Sequence 16906, A
c	25	33.8	8.4	1611	5	US-09-724-676A-16906	Sequence 16906, A
c	26	33.8	8.4	1637	5	US-09-724-676-16909	Sequence 16909, A
c	27	33.8	8.4	1637	5	US-09-724-676A-16909	Sequence 16909, A
c	28	33.8	8.4	1641	5	US-09-724-676-16904	Sequence 16904, A
c	29	33.8	8.4	1641	5	US-09-724-676A-16904	Sequence 16904, A
c	30	33.8	8.4	1667	5	US-09-724-676-16907	Sequence 16907, A
c	31	33.8	8.4	1667	5	US-09-724-676A-16907	Sequence 16907, A
c	32	33.8	8.4	2172	5	US-09-724-676-16919	Sequence 16919, A
c	33	33.8	8.4	2172	5	US-09-724-676A-16919	Sequence 16919, A
c	34	33.8	8.4	2211	5	US-09-724-676-16921	Sequence 16921, A
c	35	33.8	8.4	2211	5	US-09-724-676A-16921	Sequence 16921, A
c	36	33.8	8.4	2213	5	US-09-724-676-16917	Sequence 16917, A
c	37	33.8	8.4	2213	5	US-09-724-676A-16917	Sequence 16917, A
c	38	33.8	8.4	2237	5	US-09-724-676-16902	Sequence 16902, A
c	39	33.8	8.4	2237	5	US-09-724-676A-16902	Sequence 16902, A
c	40	33.8	8.4	2239	5	US-09-724-676-16920	Sequence 16920, A
c	41	33.8	8.4	2239	5	US-09-724-676A-16920	Sequence 16920, A
c	42	33.8	8.4	2269	5	US-09-724-676-16918	Sequence 16918, A
c	43	33.8	8.4	2269	5	US-09-724-676A-16918	Sequence 16918, A
c	44	33.8	8.4	2278	5	US-09-724-676-16900	Sequence 16900, A
c	45	33.8	8.4	2278	5	US-09-724-676A-16900	Sequence 16900, A

ALIGNMENTS

RESULT 1
US-09-531-113-41876/c
; Sequence 41876, Application US/09531113
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: Heck, Gregory R.
; APPLICANT: La Rosa, Thomas J.
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15761)B
; CURRENT APPLICATION NUMBER: US/09/531,113
; CURRENT FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 48629
; SEQ ID NO 41876
; LENGTH: 533
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: jC-gmro02910026c05a1
US-09-531-113-41876

Query Match 22.1%; Score 88.8; DB 5; Length 533;
Best Local Similarity 69.8%; Pred. No. 4.2e-17;
Matches 162; Conservative 0; Mismatches 67; Indels 3; Gaps 3;

```
Qy       5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
           | ||||| ||||| || | || | | ||||| | |||| |||||| ||  ||||||
Db     331 GTCGGGCACACCAACTCGTTCAAGAAGGTCTTAAGTATATGTTCCAGGATAA-AGGCCTT 273

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
           | |  ||| ||||| |||||  | |  |    | | ||||| |||| |||||||||  |
Db     272 GTAACTGTATGGTCCGCACCTAATTACTGTTACCGTTGTGGAAATGTAGCTTCTATTCTT 213

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
           || ||| |||| || | ||||| |||| || ||  | ||||| || || ||||| |||||
Db     212 AGTTTCAATGAAAATATGGAAA-GAGAAGTTAAATTTTTCACTGAAACAGAGGAGAACAA 154

Qy     185 CCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGT 236
           ||||||   ||| || ||| |  ||  |||||||||| | | | || | ||
Db     153 CCAGATGAGAGGCCCCAGGAC-AGGCGTCCCATATTTCTTATAAATTGATGT 103
```

RESULT 2
PCT-US02-28445-13
; Sequence 13, Application PC/TUS0228445
; GENERAL INFORMATION:
; APPLICANT: BASF PLANT SCIENCE GMBH
; TITLE OF INVENTION: PROTEIN PHOSPHATASE STRESS-RELATED POLYPEPTIDES AND
; TITLE OF INVENTION: METHODS OF USE IN PLANTS
; FILE REFERENCE: 16313-0160

```
;  CURRENT APPLICATION NUMBER: PCT/US02/28445
;  CURRENT FILING DATE:  2002-09-05
;  PRIOR APPLICATION NUMBER: 60/317,305
;  PRIOR FILING DATE: 2001-09-05
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
;   LENGTH: 1184
;   TYPE: DNA
;   ORGANISM: Glycine max
PCT-US02-28445-13

  Query Match            21.9%;  Score 88;  DB 1;  Length 1184;
  Best Local Similarity   66.2%;  Pred. No. 9.1e-17;
  Matches  172;  Conservative    0;  Mismatches   85;  Indels    3;  Gaps    3;

Qy      5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
          | || || ||||| || |  |||  | | |||||  | |||| ||| || ||  ||||||
Db    788 GTCGAGCGCACCAACTCGTTCAGGAAGGCCTTAAGTATATGTTCCAAGATAA-AGGCCTT 846

Qy     65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
          | |  ||| ||||||||||  | |  |    | ||||||| |||| |||||||||  |
Db    847 GTAACTGTATGGTCTGCACCTAATTACTGTTACCGATGTGGAAATGTAGCTTCTATTCTG 906

Qy    125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
          || ||| |||| || | ||||| |||| || ||  | ||||| || || ||||| |||||
Db    907 AGTTTCAATGAAAATATGGAAA-GAGAAGTTAAATTTTTCACCGAAACAGAGGAGAACAA 965

Qy    185 CCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGACAATT 244
          |||||   |||||| ||| |  ||  | |||||||| | | | ||   |   |   |
Db    966 TCAGATGAGAGGGCCCAGGAC-AGGCGTTCCATATTTCTTATAAGTTGGTGCAAATTTTG 1024

Qy    245 TTAGAAGGCAATATTTAAAT 264
          ||   |    ||  | ||||
Db   1025 TTTTGAATTTATTGTAAAAT 1044


RESULT 3
US-10-236-699-13
; Sequence 13, Application US/10236699
; GENERAL INFORMATION:
;  APPLICANT: BOHNERT, HANS J.
;  APPLICANT:  CHEN, RUOYING
;  APPLICANT:  ISHITANI, MANABU
;  APPLICANT:  VAN THIELEN, NOCHA
;  APPLICANT:  COSTA E SILVA, OSWALDO DA
;  TITLE OF INVENTION: PROTEIN PHOSPHATASE STRESS-RELATED POLYPEPTIDES AND
;  TITLE OF INVENTION:  METHODS OF USE IN PLANTS
;  FILE REFERENCE: 16313-0161
;  CURRENT APPLICATION NUMBER: US/10/236,699
;  CURRENT FILING DATE:  2002-09-05
;  PRIOR APPLICATION NUMBER: 60/317,305
;  PRIOR FILING DATE: 2001-09-05
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
;   LENGTH: 1184
;   TYPE: DNA
;   ORGANISM: Glycine max
US-10-236-699-13

  Query Match            21.9%;  Score 88;  DB 6;  Length 1184;
  Best Local Similarity   66.2%;  Pred. No. 9.1e-17;
  Matches  172;  Conservative    0;  Mismatches   85;  Indels    3;  Gaps    3;

Qy      5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
          | || || ||||| || |  |||  | | |||||  | |||| ||| || ||  ||||||
Db    788 GTCGAGCGCACCAACTCGTTCAGGAAGGCCTTAAGTATATGTTCCAAGATAA-AGGCCTT 846

Qy     65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
          | |  ||| ||||||||||  | |  |    | ||||||| |||| |||||||||  |
Db    847 GTAACTGTATGGTCTGCACCTAATTACTGTTACCGATGTGGAAATGTAGCTTCTATTCTG 906

Qy    125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
          || ||| |||| || | ||||| |||| || ||  | ||||| || || ||||| |||||
Db    907 AGTTTCAATGAAAATATGGAAA-GAGAAGTTAAATTTTTCACCGAAACAGAGGAGAACAA 965

Qy    185 CCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGACAATT 244
          |||||   |||||| ||| |  ||  | |||||||| | | | ||   |   |   |
Db    966 TCAGATGAGAGGGCCCAGGAC-AGGCGTTCCATATTTCTTATAAGTTGGTGCAAATTTTG 1024

Qy    245 TTAGAAGGCAATATTTAAAT 264
          ||   |    ||  | ||||
Db   1025 TTTTGAATTTATTGTAAAAT 1044


RESULT 4
US-09-531-113-2951/c
; Sequence 2951, Application US/09531113
; GENERAL INFORMATION:
;  APPLICANT: Byrum, Joseph R.
;  APPLICANT:  Heck, Gregory R.
;  APPLICANT:  La Rosa, Thomas J.
;  TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(15761)B
;  CURRENT APPLICATION NUMBER: US/09/531,113
;  CURRENT FILING DATE:  2000-03-22
;  NUMBER OF SEQ ID NOS: 48629
; SEQ ID NO 2951
;   LENGTH: 539
;   TYPE: DNA
;   ORGANISM: Glycine max
;   OTHER INFORMATION: Clone ID: jC-gmfl02220053a07d1
US-09-531-113-2951

  Query Match            21.3%;  Score 85.6;  DB 5;  Length 539;
  Best Local Similarity   63.1%;  Pred. No. 3.9e-16;
  Matches  164;  Conservative    0;  Mismatches   94;  Indels    2;  Gaps    2;

Qy      5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
          | || || ||||| || |  |||  | | |||||  | |||| ||| || ||  ||||||
Db    384 GTCGAGCGCACCAACTCGTTCAGGAAGGCCTTAAGTATATGTTCCAAGATAA-AGGCCTT 326

Qy     65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
          | |  ||| |||| ||||||  | |  |    | ||||||| |||| |||||||||  |
Db    325 GTAACTGTATGGTATGCACCTAATTACTGTTACCGATGTGGAAATGTAGCTTCTATTCTG 266

Qy    125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
          || ||| |||| || | ||||| |||| || ||  | ||||| || || ||||| |||||
Db    265 AGTTTCAATGAAAATATGGAAA-GAGAAGTTAAATTTTTCACCGAAACAGAGGAGAACAA 207

Qy    185 CCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGACAATT 244
          |||||   |||||| ||| |  |   |||    ||    |   |   |         |
Db    206 TCAGATGAGAGGGCCCAGGACAGGCGTTCCATATTTCTTATAAGATGGTGCAAATTTTGT 147

Qy    245 TTAGAAGGCAATATTTAAAT 264
          ||  ||  ||||  | ||||
Db    146 TTTTAAATCAATGGTAAAAT 127


RESULT 5
PCT-US02-28445-5
; Sequence 5, Application PC/TUS0228445
; GENERAL INFORMATION:
;  APPLICANT: BASF PLANT SCIENCE GMBH
;  TITLE OF INVENTION: PROTEIN PHOSPHATASE STRESS-RELATED POLYPEPTIDES AND
;  TITLE OF INVENTION:  METHODS OF USE IN PLANTS
;  FILE REFERENCE: 16313-0160
;  CURRENT APPLICATION NUMBER: PCT/US02/28445
;  CURRENT FILING DATE:  2002-09-05
;  PRIOR APPLICATION NUMBER: 60/317,305
;  PRIOR FILING DATE: 2001-09-05
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
```

```
;   LENGTH: 1014
;   TYPE: DNA
;   ORGANISM: Physcomitrella patens
PCT-US02-28445-5

  Query Match             15.2%;  Score 61.2;  DB 1;  Length 1014;
  Best Local Similarity   61.9%;  Pred. No. 9.6e-09;
  Matches  130;  Conservative    0;  Mismatches   78;  Indels    2;  Gaps    2;

Qy       5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
           |||| || || ||  | |  ||      |  | ||  ||||||  |  || ||  || |||
Db     730 GCCGTGCGCATCAATTAGTTCAAGAGGGATTGAAGTACATGTTTCCTGACAA-AGGACTT 788

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
           |     |||||||| || ||  | |  |    ||||||||| ||||  ||||| || |||
Db     789 GTCACGGTGTGGTCCGCTCCAAACTATTGCTACAGATGTGGAAATGTTGCTTCAATCTTA 848

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
           |||||| | || || | ||| | ||||||| ||  | || || ||||||||||| ||| |
Db     849 AGCTTCAACGAAAATATGGAGA-GAGATGTGAAATTTTTTACTGAGACCGAGGAGAACCA 907

Qy     185 CCAGATTCCAGGGCCAAGGGCTTGGAATCC 214
             ||    |  ||  | ||  |   |||
Db     908 GGCTATGATGGCACCTCGAGCAGGAGTTCC 937


RESULT 6
US-10-236-699-5
; Sequence 5, Application US/10236699
; GENERAL INFORMATION:
;  APPLICANT: BOHNERT, HANS J.
;  APPLICANT:  CHEN, RUOYING
;  APPLICANT:  ISHITANI, MANABU
;  APPLICANT:  VAN THIELEN, NOCHA
;  APPLICANT:  COSTA E SILVA, OSWALDO DA
;  TITLE OF INVENTION: PROTEIN PHOSPHATASE STRESS-RELATED POLYPEPTIDES AND
;  TITLE OF INVENTION:  METHODS OF USE IN PLANTS
;  FILE REFERENCE: 16313-0161
;  CURRENT APPLICATION NUMBER: US/10/236,699
;  CURRENT FILING DATE:  2002-09-05
;  PRIOR APPLICATION NUMBER: 60/317,305
;  PRIOR FILING DATE: 2001-09-05
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 1014
;   TYPE: DNA
;   ORGANISM: Physcomitrella patens
US-10-236-699-5

  Query Match             15.2%;  Score 61.2;  DB 6;  Length 1014;
  Best Local Similarity   61.9%;  Pred. No. 9.6e-09;
  Matches  130;  Conservative    0;  Mismatches   78;  Indels    2;  Gaps    2;

Qy       5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
           |||| || || ||  | |  ||      |  | ||  ||||||  |  || ||  || |||
Db     730 GCCGTGCGCATCAATTAGTTCAAGAGGGATTGAAGTACATGTTTCCTGACAA-AGGACTT 788

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
           |     |||||||| || ||  | |  |    ||||||||| ||||  ||||| || |||
Db     789 GTCACGGTGTGGTCCGCTCCAAACTATTGCTACAGATGTGGAAATGTTGCTTCAATCTTA 848

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
           |||||| | || || | ||| | ||||||| ||  | || || ||||||||||| ||| |
Db     849 AGCTTCAACGAAAATATGGAGA-GAGATGTGAAATTTTTTACTGAGACCGAGGAGAACCA 907

Qy     185 CCAGATTCCAGGGCCAAGGGCTTGGAATCC 214
             ||    |  ||  | ||  |   |||
Db     908 GGCTATGATGGCACCTCGAGCAGGAGTTCC 937


RESULT 7
US-10-092-411A-981/c
; Sequence 981, Application US/10092411A
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 032796-101
;  CURRENT APPLICATION NUMBER: US/10/092,411A
;  CURRENT FILING DATE:  2002-03-07
;  PRIOR APPLICATION NUMBER: US 09/134,001
;  PRIOR FILING DATE: 1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5676
; SEQ ID NO 981
;   LENGTH: 183
;   TYPE: DNA
;   ORGANISM: Staphylococcus epidermidis
US-10-092-411A-981

  Query Match              8.4%;  Score 33.8;  DB 6;  Length 183;
  Best Local Similarity   58.4%;  Pred. No. 1;
  Matches   59;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy     256 TATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGGAGGTTCAGATC 315
           | |||||||||| | |    |      |  || |||| |  |||||| |       |||  |
Db     175 TTTTTAAATCTGAACGAACTTCGTTATCATGAGATTGATTTAATTTTTCTTCATCATTTT 116

Qy     316 ATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAA 356
             ||  ||||| | |||| | ||||    ||||||| ||
Db     115 CCTGTGAAATCTGTTCCATTTCTTTTCCTAATTCTTTTAAA 75


RESULT 8
US-09-724-676-16911/c
; Sequence 16911, Application US/09724676
; GENERAL INFORMATION:
;  APPLICANT: Compugen LTD
;  TITLE OF INVENTION: Variants of alternative splicing
;  FILE REFERENCE: 129181.4 Compugen
;  CURRENT APPLICATION NUMBER: US/09/724,676
;  CURRENT FILING DATE:  2000-11-28
;  NUMBER OF SEQ ID NOS: 97222
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 16911
;   LENGTH: 1291
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (31)..(31)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (34)..(34)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (41)..(41)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (46)..(46)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (51)..(51)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (56)..(56)
;   OTHER INFORMATION: n is a,c,g, or t
```

```
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (60)..(60)
;   OTHER INFORMATION: n is a,c,g, or t
US-09-724-676-16911

  Query Match             8.4%;  Score 33.8;  DB 5;  Length 1291;
  Best Local Similarity   52.5%;  Pred. No. 1.7;
  Matches   74;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy     254 AATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGGAGGTTCAGA 313
           | | ||||  |       ||||     ||||| |    || |||||||| || |   |
Db    1262 ATTTTTTAGCTGCTTTTTGGGTTTTATACAATGAACATGTATTAATTGTAGAAGAAAACG 1203

Qy     314 TCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACAAGTACAAAAAACC 373
             | || |  | ||||  | ||||||||   |||   ||    | ||  |||||||
Db    1202 ATGTCATCCTTTATGATAAAATCCATTTCCATTTTAGCTTTTTTAAAAAAACAAAAAGCT 1143

Qy     374 AACGGGAGGGGCATAACTTCC 394
             | |    |   ||| |||
Db    1142 GTTGTGGACAGATGAACATCC 1122


RESULT 9
US-09-724-676A-16911/c
; Sequence 16911, Application US/09724676A
; GENERAL INFORMATION:
;  APPLICANT: Compugen LTD
;  TITLE OF INVENTION: Variants of alternative splicing
;  FILE REFERENCE: 129181.4 Compugen
;  CURRENT APPLICATION NUMBER: US/09/724,676A
;  CURRENT FILING DATE:  2000-11-28
;  NUMBER OF SEQ ID NOS: 97222
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 16911
;   LENGTH: 1291
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (31)..(31)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (34)..(34)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (41)..(41)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (46)..(46)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (51)..(51)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (56)..(56)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (60)..(60)
;   OTHER INFORMATION: n is a,c,g, or t
US-09-724-676A-16911

  Query Match             8.4%;  Score 33.8;  DB 5;  Length 1291;
  Best Local Similarity   52.5%;  Pred. No. 1.7;
  Matches   74;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy     254 AATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGGAGGTTCAGA 313
           | | ||||  |       ||||     ||||| |    || |||||||| || |   |
Db    1262 ATTTTTTAGCTGCTTTTTGGGTTTTATACAATGAACATGTATTAATTGTAGAAGAAAACG 1203

Qy     314 TCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACAAGTACAAAAAACC 373
             | || |  | ||||  | ||||||||   |||   ||    | ||  |||||||
Db    1202 ATGTCATCCTTTATGATAAAATCCATTTCCATTTTAGCTTTTTTAAAAAAACAAAAAGCT 1143

Qy     374 AACGGGAGGGGCATAACTTCC 394
             | |    |   ||| |||
Db    1142 GTTGTGGACAGATGAACATCC 1122


RESULT 10
US-09-724-676-16914/c
; Sequence 16914, Application US/09724676
; GENERAL INFORMATION:
;  APPLICANT: Compugen LTD
;  TITLE OF INVENTION: Variants of alternative splicing
;  FILE REFERENCE: 129181.4 Compugen
;  CURRENT APPLICATION NUMBER: US/09/724,676
;  CURRENT FILING DATE:  2000-11-28
;  NUMBER OF SEQ ID NOS: 97222
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 16914
;   LENGTH: 1317
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (31)..(31)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (34)..(34)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (41)..(41)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (46)..(46)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (51)..(51)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (56)..(56)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (60)..(60)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (182)..(182)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (188)..(188)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (198)..(199)
;   OTHER INFORMATION: n is a,c,g, or t
US-09-724-676-16914

  Query Match             8.4%;  Score 33.8;  DB 5;  Length 1317;
  Best Local Similarity   52.5%;  Pred. No. 1.7;
  Matches   74;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;
```

```
Qy     254 AATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGGAGGTTCAGA 313
           | | ||||  |       ||||     ||||| |    || |||||||| || |   |
Db    1288 ATTTTTTAGCTGCTTTTTGGGTTTTATACAATGAACATGTATTAATTGTAGAAGAAAACG 1229

Qy     314 TCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACAAGTACAAAAAACC 373
             | || |  | ||||  | ||||||||   |||   ||    | ||  ||||||| |
Db    1228 ATGTCATCCTTTATGATAAAATCCATTTCCATTTTAGCTTTTTTAAAAAAACAAAAAGCT 1169

Qy     374 AACGGGAGGGGCATAACTTCC 394
             | |    |   ||| |||
Db    1168 GTTGTGGACAGATGAACATCC 1148


RESULT 11
US-09-724-676A-16914/c
; Sequence 16914, Application US/09724676A
; GENERAL INFORMATION:
;  APPLICANT: Compugen LTD
;  TITLE OF INVENTION: Variants of alternative splicing
;  FILE REFERENCE: 129181.4 Compugen
;  CURRENT APPLICATION NUMBER: US/09/724,676A
;  CURRENT FILING DATE:  2000-11-28
;  NUMBER OF SEQ ID NOS: 97222
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 16914
;   LENGTH: 1317
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (31)..(31)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (34)..(34)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (41)..(41)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (46)..(46)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (51)..(51)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (56)..(56)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (60)..(60)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (182)..(182)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (188)..(188)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (198)..(199)
;   OTHER INFORMATION: n is a,c,g, or t
US-09-724-676A-16914

  Query Match             8.4%;  Score 33.8;  DB 5;  Length 1317;
  Best Local Similarity   52.5%;  Pred. No. 1.7;
  Matches   74;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy     254 AATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGGAGGTTCAGA 313
           | | ||||  |       ||||     ||||| |    || |||||||| || |   |
Db    1288 ATTTTTTAGCTGCTTTTTGGGTTTTATACAATGAACATGTATTAATTGTAGAAGAAAACG 1229

Qy     314 TCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACAAGTACAAAAAACC 373
             | || |  | ||||  | ||||||||   |||   ||    | ||  ||||||| |
Db    1228 ATGTCATCCTTTATGATAAAATCCATTTCCATTTTAGCTTTTTTAAAAAAACAAAAAGCT 1169

Qy     374 AACGGGAGGGGCATAACTTCC 394
             | |    |   ||| |||
Db    1168 GTTGTGGACAGATGAACATCC 1148


RESULT 12
US-09-724-676-16912/c
; Sequence 16912, Application US/09724676
; GENERAL INFORMATION:
;  APPLICANT: Compugen LTD
;  TITLE OF INVENTION: Variants of alternative splicing
;  FILE REFERENCE: 129181.4 Compugen
;  CURRENT APPLICATION NUMBER: US/09/724,676
;  CURRENT FILING DATE:  2000-11-28
;  NUMBER OF SEQ ID NOS: 97222
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 16912
;   LENGTH: 1358
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (31)..(31)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (34)..(34)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (41)..(41)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (46)..(46)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (51)..(51)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (56)..(56)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (60)..(60)
;   OTHER INFORMATION: n is a,c,g, or t
US-09-724-676-16912

  Query Match             8.4%;  Score 33.8;  DB 5;  Length 1358;
  Best Local Similarity   52.5%;  Pred. No. 1.7;
  Matches   74;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy     254 AATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGGAGGTTCAGA 313
           | | ||||  |       ||||     ||||| |    || |||||||| || |   |
Db    1329 ATTTTTTAGCTGCTTTTTGGGTTTTATACAATGAACATGTATTAATTGTAGAAGAAAACG 1270

Qy     314 TCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACAAGTACAAAAAACC 373
             | || |  | ||||  | ||||||||   |||   ||    | ||  ||||||| |
Db    1269 ATGTCATCCTTTATGATAAAATCCATTTCCATTTTAGCTTTTTTAAAAAAACAAAAAGCT 1210

Qy     374 AACGGGAGGGGCATAACTTCC 394
             | |    |   ||| |||
```

```
Db   1209 GTTGTGGACAGATGAACATCC 1189


RESULT 13
US-09-724-676A-16912/c
; Sequence 16912, Application US/09724676A
; GENERAL INFORMATION:
;  APPLICANT: Compugen LTD
;  TITLE OF INVENTION: Variants of alternative splicing
;  FILE REFERENCE: 129181.4 Compugen
;  CURRENT APPLICATION NUMBER: US/09/724,676A
;  CURRENT FILING DATE:  2000-11-28
;  NUMBER OF SEQ ID NOS: 97222
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 16912
;   LENGTH: 1358
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (31)..(31)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (34)..(34)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (41)..(41)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (46)..(46)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (51)..(51)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (56)..(56)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (60)..(60)
;   OTHER INFORMATION: n is a,c,g, or t
US-09-724-676A-16912

  Query Match             8.4%;  Score 33.8;  DB 5;  Length 1358;
  Best Local Similarity   52.5%;  Pred. No. 1.7;
  Matches   74;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy    254 AATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGGAGGTTCAGA 313
          | | ||||  |       ||||     ||||| |    || |||||||| || |   |
Db   1329 ATTTTTTAGCTGCTTTTTGGGTTTTATACAATGAACATGTATTAATTGTAGAAGAAAACG 1270

Qy    314 TCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACAAGTACAAAAAACC 373
            | || |  | ||||  | ||||||||   |||   ||    | ||  ||||||| |
Db   1269 ATGTCATCCTTTATGATAAAATCCATTTCCATTTTAGCTTTTTTAAAAAAACAAAAAGCT 1210

Qy    374 AACGGGAGGGGCATAACTTCC 394
            | |    |   ||| |||
Db   1209 GTTGTGGACAGATGAACATCC 1189


RESULT 14
US-09-724-676-16915/c
; Sequence 16915, Application US/09724676
; GENERAL INFORMATION:
;  APPLICANT: Compugen LTD
;  TITLE OF INVENTION: Variants of alternative splicing
;  FILE REFERENCE: 129181.4 Compugen
;  CURRENT APPLICATION NUMBER: US/09/724,676
;  CURRENT FILING DATE:  2000-11-28
;  NUMBER OF SEQ ID NOS: 97222
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 16915
;   LENGTH: 1384
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (31)..(31)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (34)..(34)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (41)..(41)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (46)..(46)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (51)..(51)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (56)..(56)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (60)..(60)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (182)..(182)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (188)..(188)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (198)..(199)
;   OTHER INFORMATION: n is a,c,g, or t
US-09-724-676-16915

  Query Match             8.4%;  Score 33.8;  DB 5;  Length 1384;
  Best Local Similarity   52.5%;  Pred. No. 1.7;
  Matches   74;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy    254 AATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGGAGGTTCAGA 313
          | | ||||  |       ||||     ||||| |    || |||||||| || |   |
Db   1355 ATTTTTTAGCTGCTTTTTGGGTTTTATACAATGAACATGTATTAATTGTAGAAGAAAACG 1296

Qy    314 TCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACAAGTACAAAAAACC 373
            | || |  | ||||  | ||||||||   |||   ||    | ||  ||||||| |
Db   1295 ATGTCATCCTTTATGATAAAATCCATTTCCATTTTAGCTTTTTTAAAAAAACAAAAAGCT 1236

Qy    374 AACGGGAGGGGCATAACTTCC 394
            | |    |   ||| |||
Db   1235 GTTGTGGACAGATGAACATCC 1215


RESULT 15
US-09-724-676A-16915/c
; Sequence 16915, Application US/09724676A
; GENERAL INFORMATION:
;  APPLICANT: Compugen LTD
;  TITLE OF INVENTION: Variants of alternative splicing
;  FILE REFERENCE: 129181.4 Compugen
```

```
;  CURRENT APPLICATION NUMBER: US/09/724,676A
;  CURRENT FILING DATE:  2000-11-28
;  NUMBER OF SEQ ID NOS: 97222
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 16915
;   LENGTH: 1384
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (31)..(31)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (34)..(34)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (41)..(41)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (46)..(46)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (51)..(51)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (56)..(56)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (60)..(60)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (182)..(182)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (188)..(188)
;   OTHER INFORMATION: n is a,c,g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (198)..(199)
;   OTHER INFORMATION: n is a,c,g, or t
US-09-724-676A-16915

  Query Match             8.4%;  Score 33.8;  DB 5;  Length 1384;
  Best Local Similarity   52.5%;  Pred. No. 1.7;
  Matches   74;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy     254 AATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGGAGGTTCAGA 313
           | | ||||  |       ||||     ||||| |    || |||||||| || |   |
Db    1355 ATTTTTTAGCTGCTTTTTGGGTTTTATACAATGAACATGTATTAATTGTAGAAGAAAACG 1296

Qy     314 TCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACAAGTACAAAAAACC 373
           | || |  | ||||  | ||||||||   |||   ||    | ||  ||||||| |
Db    1295 ATGTCATCCTTTATGATAAAATCCATTTCCATTTTAGCTTTTTTAAAAAAACAAAAAGCT 1236

Qy     374 AACGGGAGGGGCATAACTTCC 394
             | |    |   ||| |||
Db    1235 GTTGTGGACAGATGAACATCC 1215


Search completed: January  2, 2003, 15:04:23
Job time : 149 secs
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: January 2, 2003, 12:30:02 ; Search time 2188 Seconds
(without alignments)
2975.588 Million cell updates/sec

Title: US-09-850-147-1
Perfect score: 402
Sequence: 1 ccgggccgggctcaccagct.....ggcataacttcctgtgttcc 402

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	266	66.2	568	10	AW565695	AW565695 LG1_348_A
2	266	66.2	606	14	BQ656076	BQ656076 pRC0407 S
3	266	66.2	612	10	AW745589	AW745589 WS1_35_A0
4	266	66.2	614	10	BE597103	BE597103 PI1_70_G0
5	266	66.2	681	13	BG946826	BG946826 IP1_1_B10
6	252.4	62.8	493	12	BG558648	BG558648 RHIZ2_58_
c 7	229.4	57.1	479	10	AW091358	AW091358 614094A10
c 8	229.4	57.1	570	9	AI978092	AI978092 614022D04
9	229.4	57.1	592	10	AW036911	AW036911 614022D04
c 10	227.8	56.7	1349	11	AY109457	AY109457 Zea mays
c 11	221.8	55.2	470	9	AI621443	AI621443 486090D01
c 12	220	54.7	590	10	AW288508	AW288508 618068C02
c 13	220	54.7	613	9	AI691590	AI691590 606022D02
c 14	218	54.2	413	10	AW924010	AW924010 WS1_32_C0
c 15	212.4	52.8	605	13	BM079618	BM079618 MEST97-B0
c 16	210.4	52.3	537	12	BG842699	BG842699 MEST39-A0
c 17	195	48.5	365	13	BI388689	BI388689 949044B01
18	195	48.5	400	13	BI396083	BI396083 949044B01
c 19	186.4	46.4	352	12	BG354266	BG354266 947033G03
20	181	45.0	487	10	BE598350	BE598350 PI1_81_E1
c 21	178.6	44.4	604	13	BJ284678	BJ284678 BJ284678
c 22	175.8	43.7	301	9	AI902083	AI902083 614022D04
c 23	173.8	43.2	634	13	BJ251011	BJ251011 BJ251011
24	167.4	41.6	535	10	AW679158	AW679158 WS1_22_A0
25	166.6	41.4	414	9	AU031687	AU031687 AU031687
c 26	154.8	38.5	570	13	BJ466432	BJ466432 BJ466432
c 27	153.2	38.1	356	9	AI619388	AI619388 486093F06
c 28	152.6	38.0	492	10	AV943812	AV943812 AV943812
c 29	152.6	38.0	653	13	BJ482869	BJ482869 BJ482869
c 30	152.4	37.9	665	13	BJ211965	BJ211965 BJ211965
31	151	37.6	536	9	AJ435241	AJ435241 AJ435241
c 32	151	37.6	547	14	BQ664779	BQ664779 HV04N10u
33	149.2	37.1	578	10	AW065323	AW065323 614045H08
c 34	148.8	37.0	511	9	AI979736	AI979736 614045H08
c 35	148.6	37.0	593	9	AL499918	AL499918 AL499918
c 36	140.4	34.9	587	13	BJ486368	BJ486368 BJ486368
c 37	138.4	34.4	531	12	BG265579	BG265579 1000029E0
c 38	116	28.9	693	13	BJ322756	BJ322756 BJ322756
39	112.6	28.0	630	14	BQ472249	BQ472249 HV04N10r
c 40	111	27.6	707	10	AV946097	AV946097 AV946097
c 41	108.8	27.1	610	9	AL500634	AL500634 AL500634
42	107.2	26.7	477	12	BF199841	BF199841 WHE0597_B
c 43	93.6	23.3	624	14	BQ510007	BQ510007 EST617422
44	93	23.1	539	10	AW041153	AW041153 EST284017
c 45	92	22.9	481	14	BQ120114	BQ120114 EST605690

ALIGNMENTS

RESULT 1
AW565695
LOCUS AW565695 568 bp mRNA linear EST 19-JUL-2000
DEFINITION LG1_348_A05.g1_A002 Light Grown 1 (LG1) Sorghum bicolor cDNA, mRNA sequence.
ACCESSION AW565695
VERSION AW565695.1 GI:7219573
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 568)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C. and Pratt,L.H.
TITLE An EST database from Sorghum: light-grown seedlings
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: T7
High quality sequence start: 29

High quality sequence stop: 568
POLYA=Yes.

```
FEATURES             Location/Qualifiers
     source          1. .568
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Light Grown 1 (LG1)"
                     /note="Organ: 10- to 14-day-old light-grown (greenhouse)
                     seedlings; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI
                     ; The library was made from poly-A RNA in the cloning
                     vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT      147 a     99 c    142 g    180 t
ORIGIN
```

Query Match 66.2%; Score 266; DB 10; Length 568;
Best Local Similarity 84.2%; Pred. No. 6.2e-71;
Matches 335; Conservative 0; Mismatches 60; Indels 3; Gaps 3;

```
Qy       5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
Db     119 GCCGGGCTCACCAGCTGGTCCAGGAAGGCTTAAAGTACATGTTCCAGGAGAA-GGGCCTT 177

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
Db     178 GTAACTGTGTGGTCTGCACCTAATTATTGCTACAGATGTGGCAATGTAGCTTCTATATTA 237

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
Db     238 AGCTTCGATGAGAAGATGGAAA-GAGATGTCAAGTTCTTCACAGAGACAGAGGAGAACAA 296

Qy     185 CCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGACAATT 244
Db     297 CCAGATGCGAGGCCCAAGGAC-TGCAGTCCCATATTTCCTCTGACTTCAGTTGTACAATT 355

Qy     245 TTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGG 304
Db     356 TTAGAAGTCAATATTTAAATCTGTAAGGGTTAAAGATCAATCAGATGTGTTTATTGTTGT 415

Qy     305 AGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACAAGTA 364
Db     416 AGGTTCAGATCATAGTGCAATCTGATGCATTCGATTTCTGTTTTCTTTTGACCACATGTA 475

Qy     365 CAAAAAACCAACGGGAGGGGCATAACTTCCTGTGTTCC 402
Db     476 CAATAGACCATCGGGAGTGTCATATCTTCCTGTATTCC 513
```

```
RESULT 2
BQ656076
LOCUS       BQ656076                606 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  pRC0407 Sorghum halepense rhizome cDNA library Sorghum halepense
            cDNA clone pRC0407, mRNA sequence.
ACCESSION   BQ656076
VERSION     BQ656076.1  GI:21788402
KEYWORDS    EST.
SOURCE      Sorghum halepense.
  ORGANISM  Sorghum halepense
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 606)
  AUTHORS   Paterson,A.H.
  TITLE     Unpublished, Paterson,A.H
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Paterson AH
            Center for Applied Genetic Technologies
            University of Georgia
            Riverbend Research Laboratory, Room 162, 110 Riverbend Road, Athens
            , GA 30602, USA
            Tel: 7065830162
            Fax: 7065830160
            Email: paterson@uga.edu
```

contig
Seq primer: M13F/M13R.

```
FEATURES             Location/Qualifiers
     source          1. .606
                     /organism="Sorghum halepense"
                     /db_xref="taxon:4560"
                     /clone="pRC0407"
                     /clone_lib="Sorghum halepense rhizome cDNA library"
                     /tissue_type="rhizome"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                     XhoI; cDNA library was prepared from mRNA isolated from
                     the apical 2-3 cm of actively growing rhizomes of Johnson
                     grass (Sorghum halepense) and cloned into the EcoRI/XhoI
                     sites of lambda ZAP II (Stratagene)."
BASE COUNT      152 a    110 c    147 g    197 t
ORIGIN
```

Query Match 66.2%; Score 266; DB 14; Length 606;
Best Local Similarity 84.2%; Pred. No. 6.3e-71;
Matches 335; Conservative 0; Mismatches 60; Indels 3; Gaps 3;

```
Qy       5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
Db      90 GCCGGGCTCACCAGCTGGTCCAGGAAGGCTTAAAGTACATGTTCCAGGAGAA-GGGCCTT 148

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
Db     149 GTAACTGTGTGGTCTGCACCTAATTATTGCTACAGATGTGGCAATGTAGCTTCTATATTA 208

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
Db     209 AGCTTCGATGAGAAGATGGAAA-GAGATGTCAAGTTCTTCACAGAGACAGAGGAGAACAA 267

Qy     185 CCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGACAATT 244
Db     268 CCAGATGCGAGGCCCAAGGAC-TGCAGTCCCATATTTCCTCTGACTTCAGTTGTACAATT 326

Qy     245 TTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGG 304
Db     327 TTAGAAGTCAATATTTAAATCTGTAAGGGTTAAAGATCAATCAGATGTGTTTATTGTTGT 386

Qy     305 AGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACAAGTA 364
Db     387 AGGTTCAGATCATAGTGCAATCTGATGCATTCGATTTCTGTTTTCTTTTGACCACATGTA 446

Qy     365 CAAAAAACCAACGGGAGGGGCATAACTTCCTGTGTTCC 402
Db     447 CAATAGACCATCGGGAGTGTCATATCTTCCTGTATTCC 484
```

```
RESULT 3
AW745589
LOCUS       AW745589                612 bp    mRNA    linear   EST 19-JUL-2000
DEFINITION  WS1_35_A09.g1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   AW745589
VERSION     AW745589.1  GI:7659327
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 612)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: water-stressed plants
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
```

```
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: T7
            High quality sequence start: 7
            High quality sequence stop: 612
            POLYA=Yes.
FEATURES              Location/Qualifiers
     source           1. .612
                      /organism="Sorghum bicolor"
                      /db_xref="taxon:4558"
                      /clone_lib="Water-stressed 1 (WS1)"
                      /note="Organ: Mix of 5-week old plants on days 7 & 8 after
                      water was withheld; Vector: Lambda Zap; Site_1: XhoI;
                      Site_2: EcoRI; The library was made from poly-A RNA in the
                      cloning vector lambda ZAP II.  Clones to be sequenced were
                      prepared by mass excision."
BASE COUNT      156 a    110 c    148 g    198 t
ORIGIN

  Query Match             66.2%;  Score 266;  DB 10;  Length 612;
  Best Local Similarity   84.2%;  Pred. No. 6.3e-71;
  Matches  335;  Conservative    0;  Mismatches   60;  Indels    3;  Gaps    3;

Qy       5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
          ||||||||||||||||||| ||||  | |  | ||  |||||| |||||||| |||||||
Db      97 GCCGGGCTCACCAGCTGGTCCAGGAAGGCTTAAAGTACATGTTCCAGGAGAA-GGGCCTT 155

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
          | |  ||||||||||||||  | |  |    |||||||||||||| ||||||||||||
Db     156 GTAACTGTGTGGTCTGCACCTAATTATTGCTACAGATGTGGCAATGTAGCTTCTATATTA 215

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
          |||||||||||||||| ||||| |||||||||| |||||||||||| ||||| |||||
Db     216 AGCTTCGATGAGAAGATGGAAA-GAGATGTCAAGTTCTTCACAGAGACAGAGGAGAACAA 274

Qy     185 CCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGACAATT 244
          |||||| | ||| |||||| | || | |||||||||||||| ||||||||| || ||||||
Db     275 CCAGATGCGAGGCCCAAGGAC-TGCAGTCCCATATTTCCTCTGACTTCAGTTGTACAATT 333

Qy     245 TTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGG 304
          ||||||| ||||||||||||||| ||||| |||||| ||||||||||   ||| |  ||
Db     334 TTAGAAGTCAATATTTAAATCTGTAAGGGTTAAAGATCAATCAGATGTGTTTATTGTTGT 393

Qy     305 AGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACAAGTA 364
          |||||||||||||| ||| |||||||||||||| ||||||| |||||||||| |||| |||
Db     394 AGGTTCAGATCATAGTGCAATCTGATGCATTCGATTTCTGTTTTCTTTTGACCACATGTA 453

Qy     365 CAAAAAACCAACGGGAGGGGCATAACTTCCTGTGTTCC 402
          ||| | |||| |||||| | |||| |||||||| ||||
Db     454 CAATAGACCATCGGGAGTGTCATATCTTCCTGTATTCC 491


RESULT 4
BE597103
LOCUS       BE597103                 614 bp    mRNA    linear   EST 18-AUG-2000
DEFINITION  PI1_70_G02.g1_A002 Pathogen induced 1 (PI1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   BE597103
VERSION     BE597103.1  GI:9852288
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 614)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: pathogen-induced plants
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: PolyTMix
            High quality sequence start: 12
            High quality sequence stop: 604
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .614
                      /organism="Sorghum bicolor"
                      /db_xref="taxon:4558"
                      /clone_lib="Pathogen induced 1 (PI1)"
                      /note="Organ: Anthracnose-infected leaves from
                      two-week-old sorghum plants 48 hr after inoculation;
                      Vector: pBluescript II from Lambda Zap II; Site_1: XhoI;
                      Site_2: EcoRI; Two-week-old sorghum plants (BTX 623
                      cultivar) were infected with pathogen (isolate FRM42I of
                      Colletotrichum graminicola, which is a sorghum isolate).
                      RNA was prepared from infected leaves harvested from 45
                      seedlings  48 hours after inoculation.  Note: young
                      seedlings (2 weeks old) exhibit juvenile resistant
                      reaction, which is an incompatible interaction. As they
                      grow older (4 weeks or older), plants resume susceptiblity
                      to anthracnose disease. The library was made from poly-A
                      RNA in the cloning vector lambda ZAP II. Clones to be
                      sequenced were prepared by mass excision. WARNING: While
                      most or all ESTs are expected to derive from the host
                      plant, no effort was made to eliminate ESTs deriving from
                      the pathogen."
BASE COUNT      156 a    110 c    154 g    193 t      1 others
ORIGIN

  Query Match             66.2%;  Score 266;  DB 10;  Length 614;
  Best Local Similarity   84.2%;  Pred. No. 6.4e-71;
  Matches  335;  Conservative    0;  Mismatches   60;  Indels    3;  Gaps    3;

Qy       5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
          ||||||||||||||||||| ||||  | |  | ||  |||||| |||||||| |||||||
Db     123 GCCGGGCTCACCAGCTGGTCCAGGAAGGCTTAAAGTACATGTTCCAGGAGAA-GGGCCTT 181

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
          | |  ||||||||||||||  | |  |    |||||||||||||| ||||||||||||
Db     182 GTAACTGTGTGGTCTGCACCTAATTATTGCTACAGATGTGGCAATGTAGCTTCTATATTA 241

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
          |||||||||||||||| ||||| |||||||||| |||||||||||| ||||| |||||
Db     242 AGCTTCGATGAGAAGATGGAAA-GAGATGTCAAGTTCTTCACAGAGACAGAGGAGAACAA 300

Qy     185 CCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGACAATT 244
          |||||| | ||| |||||| | || | |||||||||||||| ||||||||| || ||||||
Db     301 CCAGATGCGAGGCCCAAGGAC-TGCAGTCCCATATTTCCTCTGACTTCAGTTGTACAATT 359

Qy     245 TTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGG 304
          ||||||| ||||||||||||||| ||||| |||||| ||||||||||   ||| |  ||
Db     360 TTAGAAGTCAATATTTAAATCTGTAAGGGTTAAAGATCAATCAGATGTGTTTATTGTTGT 419

Qy     305 AGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACAAGTA 364
          |||||||||||||| ||| |||||||||||||| ||||||| |||||||||| |||| |||
Db     420 AGGTTCAGATCATAGTGCAATCTGATGCATTCGATTTCTGTTTTCTTTTGACCACATGTA 479

Qy     365 CAAAAAACCAACGGGAGGGGCATAACTTCCTGTGTTCC 402
          ||| | |||| |||||| | |||| |||||||| ||||
Db     480 CAATAGACCATCGGGAGTGTCATATCTTCCTGTATTCC 517


RESULT 5
```

```
BG946826
LOCUS       BG946826                681 bp    mRNA    linear   EST 12-JUN-2001
DEFINITION  IP1_1_B10.g1_A002 Immature pannicle 1 (IP1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   BG946826
VERSION     BG946826.1  GI:14365015
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 681)
  AUTHORS   Klein,R.R., Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: developing preanthesis pannicles
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for high quality sequence is
            20. Three-prime sequences, which are obtained with PolyTMix or T7
            sequencing primer, are presented as the reverse complement.
            Seq primer: T7
            High quality sequence start: 56
            High quality sequence stop: 681
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .681
                     /organism="Sorghum bicolor"
                     /cultivar="BTx623"
                     /db_xref="taxon:4558"
                     /clone_lib="Immature pannicle 1 (IP1)"
                     /note="Organ: Developing preanthesis pannicles; Vector:
                     pBluescript II SK(-) from Lambda Zap II; Site_1: XhoI;
                     Site_2: EcoRI; The library was made from poly-A RNA in the
                     cloning vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT      177 a    129 c    163 g    212 t
ORIGIN

  Query Match             66.2%;  Score 266;  DB 13;  Length 681;
  Best Local Similarity   84.2%;  Pred. No. 6.5e-71;
  Matches  335;  Conservative    0;  Mismatches   60;  Indels    3;  Gaps    3;

Qy       5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
          ||||||||||||||||||| ||||  | |  | ||  |||||| |||||||| |||||||
Db     255 GCCGGGCTCACCAGCTGGTCCAGGAAGGCTTAAAGTACATGTTCCAGGAGAA-GGGCCTT 313

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
          | |  ||||||||||||||  | |  |    |||||||||||||| ||||||||||||
Db     314 GTAACTGTGTGGTCTGCACCTAATTATTGCTACAGATGTGGCAATGTAGCTTCTATATTA 373

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
          |||||||||||||||| ||||| |||||||||| |||||||||||| ||||| |||||
Db     374 AGCTTCGATGAGAAGATGGAAA-GAGATGTCAAGTTCTTCACAGAGACAGAGGAGAACAA 432

Qy     185 CCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGACAATT 244
          |||||| | ||| |||||| | || | |||||||||||| ||||||||| || ||||||
Db     433 CCAGATGCGAGGCCCAAGGAC-TGCAGTCCCATATTTCCTCTGACTTCAGTTGTACAATT 491

Qy     245 TTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGG 304
          ||||||| |||||||||||||| ||||| |||||| ||||||||||   ||| |  ||
Db     492 TTAGAAGTCAATATTTAAATCTGTAAGGGTTAAAGATCAATCAGATGTGTTTATTGTTGT 551

Qy     305 AGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACAAGTA 364
          |||||||||||||| ||| |||||||||||||| ||||||| |||||||||| |||| |||
Db     552 AGGTTCAGATCATAGTGCAATCTGATGCATTCGATTTCTGTTTTCTTTTGACCACATGTA 611

Qy     365 CAAAAAACCAACGGGAGGGGCATAACTTCCTGTGTTCC 402
          ||| | |||| |||||| | |||| |||||||| ||||
Db     612 CAATAGACCATCGGGAGTGTCATATCTTCCTGTATTCC 649


RESULT 6
BG558648
LOCUS       BG558648                493 bp    mRNA    linear   EST 10-APR-2001
DEFINITION  RHIZ2_58_E10.g1_A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA
            sequence.
ACCESSION   BG558648
VERSION     BG558648.1  GI:13587646
KEYWORDS    EST.
SOURCE      Sorghum propinquum.
  ORGANISM  Sorghum propinquum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 493)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: Sorghum propinquum rhizomes
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Seq primer: T7
            High quality sequence start: 70
            High quality sequence stop: 419
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .493
                     /organism="Sorghum propinquum"
                     /db_xref="taxon:132711"
                     /clone_lib="Rhizome2 (RHIZ2)"
                     /note="Organ: Rhizomes; Vector: pBluescript II from Lambda
                     Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                     from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      131 a     90 c    115 g    157 t
ORIGIN

  Query Match             62.8%;  Score 252.4;  DB 12;  Length 493;
  Best Local Similarity   83.7%;  Pred. No. 9.4e-67;
  Matches  333;  Conservative    0;  Mismatches   61;  Indels    4;  Gaps    4;

Qy       5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
          ||||||||||||||||||| ||||  | |  | ||  |||||| |||||||| |||||||
Db      50 GCCGGGCTCACCAGCTGGTCCAGGAAGGCTTAAAGTACATGTTCCAGGAGAA-GGGCCTT 108

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
          | |  ||||||||||||||  | |  |    |||||||||||||| ||||||||||||
Db     109 GTAACTGTGTGGTCTGCACCTAATTATTGCTACAGATGTGGCAATGTAGCTTCTATATTA 168

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
          |||||||||||||||| ||||| |||||||||| |||||||||||| ||||| |||||
Db     169 AGCTTCGATGAGAAGATGGAAA-GAGATGTCAAGTTCTTCACAGAGACAGAGGAGAACAA 227

Qy     185 CCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGACAATT 244
          |||||| | ||| |||||| | || | |||||||||||| ||||||||| || ||||||
Db     228 CCAGATGCGAGGCCCAAGGAC-TGCAGTCCCATATTTCCTCTGACTTCAGTTGTACAATT 286

Qy     245 TTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGG 304
          ||||||| |||||||||||||| ||||| |||||| ||||||||||   ||| |  ||
Db     287 TTAGAAGTCAATATTTAAATCTGTAAGGGTTAAAGATCAATCAGATGTGTTTATTGTTGT 346

Qy     305 AGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACAAGTA 364
          |||||||||||||| ||| |||||||||||||| ||||||| |||||||||| |||| |||
```

```
Db    347 AGGTTCAGATCATAGTGCAATCTGATGCATTCGATTTCTGTTTTCTTTTGACCACATGTA 406

Qy    365 CAAAAAACCAACGGGAGGGGCATAACTTCCTGTGTTCC 402
          ||| | |||  |||||| | || | |||||||| ||||
Db    407 CAATAGACCTGCGGGAGTGTCA-ATCTTCCTGTATTCC 443


RESULT 7
AW091358/c
LOCUS       AW091358                479 bp    mRNA    linear   EST 18-OCT-1999
DEFINITION  614094A10.x1 614 - root cDNA library from Walbot Lab Zea mays cDNA,
            mRNA sequence.
ACCESSION   AW091358
VERSION     AW091358.1  GI:6056953
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 479)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 614094  row: A  column: 10.
FEATURES             Location/Qualifiers
     source          1. .479
                     /organism="Zea mays"
                     /cultivar="W23"
                     /db_xref="taxon:4577"
                     /clone_lib="614 - root cDNA library from Walbot Lab"
                     /tissue_type="root"
                     /dev_stage="3-4 days old"
                     /lab_host="XLOLR"
                     /note="Organ: root; Vector: pBlueScriptII SK+; Site_1:
                     EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot
                     Lab (LM)"
BASE COUNT      142 a    111 c     96 g    130 t
ORIGIN

  Query Match             57.1%;  Score 229.4;  DB 10;  Length 479;
  Best Local Similarity   81.7%;  Pred. No. 1.1e-59;
  Matches  313;  Conservative    0;  Mismatches   66;  Indels    4;  Gaps    4;

Qy      5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
          |||||||||||||||||| ||||  | |  | ||  |||||| |||||||| |||||||
Db    405 GCCGGGCTCACCAGCTGGTCCAGGAAGGCTTAAAGTACATGTTCCAGGAGAA-GGGCCTT 347

Qy     65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
          | |  |||||||||||||||| | |  |    |||||||||||||| ||||||||||||
Db    346 GTAACTGTGTGGTCTGCACCCAATTATTGCTACAGATGTGGCAATGTAGCTTCTATATTA 287

Qy    125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
          |||||||||||||||| ||||| |||||||||||| ||||||||||||| ||||||||||
Db    286 AGCTTCGATGAGAAGATGGAAA-GAGATGTCAAGTTCTTCACAGAGACAGAGGAAAACAA 228

Qy    185 CCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGACAATT 244
          ||||||  | ||| |||||| | || | |||||||||||||| ||||||||  || || |||
Db    227 CCAGATGCGAGGCCCAAGGAC-TGCAGTCCCATATTTCCTCTGACTTCAATTGTAC-ATT 170

Qy    245 TTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGG 304
          ||||||| ||| |||||||||||| ||||| |||||| || |||||||   |||    ||
Db    169 TTAGAAGTCAACATTTAAATCTGTAAGGGCTAAAGATCAGTCAGATGTGTTTATCATTGT 110

Qy    305 AGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACAAGTA 364
          ||||||||||||||| ||| |||| |||||||  ||||||  |||||||| | |||| |||
Db    109 AGGTTCAGATCATAGTGCAATCTTATGCATTTGATTTCTCTTTTCTTTTAACCACATGTA 50

Qy    365 CAAAAAACCAACGGGAGGGGCAT 387
          ||| | |||| | ||||   |||
Db     49 CAATAGACCATCCGGAGTCTCAT 27


RESULT 8
AI978092/c
LOCUS       AI978092                570 bp    mRNA    linear   EST 27-AUG-1999
DEFINITION  614022D04.x2 614 - root cDNA library from Walbot Lab Zea mays cDNA,
            mRNA sequence.
ACCESSION   AI978092
VERSION     AI978092.1  GI:5791300
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 614022  row: D  column: 04.
FEATURES             Location/Qualifiers
     source          1. .570
                     /organism="Zea mays"
                     /cultivar="W23"
                     /db_xref="taxon:4577"
                     /clone_lib="614 - root cDNA library from Walbot Lab"
                     /tissue_type="root"
                     /dev_stage="3-4 days old"
                     /lab_host="XLOLR"
                     /note="Organ: root; Vector: pBlueScriptII SK+; Site_1:
                     EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot
                     Lab (LM)"
BASE COUNT      177 a    133 c    112 g    148 t
ORIGIN

  Query Match             57.1%;  Score 229.4;  DB 9;  Length 570;
  Best Local Similarity   81.7%;  Pred. No. 1.2e-59;
  Matches  313;  Conservative    0;  Mismatches   66;  Indels    4;  Gaps    4;

Qy      5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
          |||||||||||||||||| ||||  | |  | ||  |||||| |||||||| |||||||
Db    497 GCCGGGCTCACCAGCTGGTCCAGGAAGGCTTAAAGTACATGTTCCAGGAGAA-GGGCCTT 439

Qy     65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
          | |  |||||||||||||||| | |  |    |||||||||||||| ||||||||||||
Db    438 GTAACTGTGTGGTCTGCACCCAATTATTGCTACAGATGTGGCAATGTAGCTTCTATATTA 379

Qy    125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
          |||||||||||||||| ||||| |||||||||||| ||||||||||||| ||||||||||
Db    378 AGCTTCGATGAGAAGATGGAAA-GAGATGTCAAGTTCTTCACAGAGACAGAGGAAAACAA 320

Qy    185 CCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGACAATT 244
          ||||||  | ||| |||||| | || | |||||||||||||| ||||||||  || || |||
Db    319 CCAGATGCGAGGCCCAAGGAC-TGCAGTCCCATATTTCCTCTGACTTCAATTGTAC-ATT 262

Qy    245 TTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGG 304
          ||||||| ||| |||||||||||| ||||| |||||| || |||||||   |||    ||
Db    261 TTAGAAGTCAACATTTAAATCTGTAAGGGCTAAAGATCAGTCAGATGTGTTTATCATTGT 202
```

```
Qy    305 AGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACAAGTA 364
          ||||||||||||||| ||| |||| ||||||| ||||||  |||||||| | |||| |||
Db    201 AGGTTCAGATCATAGTGCAATCTTATGCATTTGATTTCTCTTTTCTTTTAACCACATGTA 142

Qy    365 CAAAAAACCAACGGGAGGGGCAT 387
          ||| | |||| | ||||   |||
Db    141 CAATAGACCATCCGGAGTCTCAT 119


RESULT 9
AW036911
LOCUS       AW036911                592 bp    mRNA    linear   EST 15-SEP-1999
DEFINITION  614022D04.y1 614 - root cDNA library from Walbot Lab Zea mays cDNA,
            mRNA sequence.
ACCESSION   AW036911
VERSION     AW036911.1  GI:5895665
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 592)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 614022  row: D  column: 04.
FEATURES             Location/Qualifiers
     source          1. .592
                     /organism="Zea mays"
                     /cultivar="W23"
                     /db_xref="taxon:4577"
                     /clone_lib="614 - root cDNA library from Walbot Lab"
                     /tissue_type="root"
                     /dev_stage="3-4 days old"
                     /lab_host="XLOLR"
                     /note="Organ: root; Vector: pBlueScriptII SK+; Site_1:
                     EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot
                     Lab (LM)"
BASE COUNT      159 a    120 c    142 g    171 t
ORIGIN

  Query Match             57.1%;  Score 229.4;  DB 10;  Length 592;
  Best Local Similarity   81.7%;  Pred. No. 1.2e-59;
  Matches  313;  Conservative    0;  Mismatches   66;  Indels    4;  Gaps    4;

Qy      5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
          |||||||||||||||||| |||| | | | ||  |||||| |||||||| |||||||
Db    199 GCCGGGCTCACCAGCTGGTCCAGGAAGGCTTAAAGTACATGTTCCAGGAGAA-GGGCCTT 257

Qy     65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
          | |  ||||||||||||||||| | |  |    ||||||||||||| ||||||||||||
Db    258 GTAACTGTGTGGTCTGCACCCAATTATTGCTACAGATGTGGCAATGTAGCTTCTATATTA 317

Qy    125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
          |||||||||||||||| ||||| |||||||||| |||||||||||| ||||||||||
Db    318 AGCTTCGATGAGAAGATGGAAA-GAGATGTCAAGTTCTTCACAGAGACAGAGGAAAACAA 376

Qy    185 CCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGACAATT 244
          |||||| | ||| |||||| | || | |||||||||||||| |||||||  || || |||
Db    377 CCAGATGCGAGGCCCAAGGAC-TGCAGTCCCATATTTCCTCTGACTTCAATTGTAC-ATT 434

Qy    245 TTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGG 304
          ||||||| ||| |||||||||| ||||| |||||| || |||||||   |||    ||
Db    435 TTAGAAGTCAACATTTAAATCTGTAAGGGCTAAAGATCAGTCAGATGTGTTTATCATTGT 494

Qy    305 AGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACAAGTA 364
          |||||||||||||| ||| |||| ||||||| ||||||  |||||||| | |||| |||
Db    495 AGGTTCAGATCATAGTGCAATCTTATGCATTTGATTTCTCTTTTCTTTTAACCACATGTA 554

Qy    365 CAAAAAACCAACGGGAGGGGCAT 387
          ||| | |||| | ||||   |||
Db    555 CAATAGACCATCCGGAGTCTCAT 577


RESULT 10
AY109457/c
LOCUS       AY109457                1349 bp    mRNA    linear   HTC 25-MAY-2002
DEFINITION  Zea mays CL12356_1 mRNA sequence.
ACCESSION   AY109457
VERSION     AY109457.1  GI:21213184
KEYWORDS    HTC.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 1349)
  AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
  TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
  JOURNAL   Unpublished (2002)
REFERENCE   2  (bases 1 to 1349)
  AUTHORS   Coe,E.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
FEATURES             Location/Qualifiers
     source          1. .1349
                     /organism="Zea mays"
                     /db_xref="MaizeDB:630146"
                     /db_xref="taxon:4577"
                     /clone="CL12356_1"
                     /clone_lib="Maize Mapping Project/DuPont Cornsensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
BASE COUNT      350 a    314 c    286 g    310 t     89 others
ORIGIN

  Query Match             56.7%;  Score 227.8;  DB 11;  Length 1349;
  Best Local Similarity   81.5%;  Pred. No. 4.7e-59;
  Matches  312;  Conservative    0;  Mismatches   67;  Indels    4;  Gaps    4;

Qy      5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
          |||||||||||||||||| |||| | |  | ||  |||||| |||||||| |||||||
Db    496 GCCGGGCTCACCAGCTGGTCCAGGAAGGCTTAAAGTACATGTTCCAGGAGAA-GGGCCTT 438

Qy     65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
          | |  ||||||||||||||||| | |  |    ||||||||||||| ||||||||||||
Db    437 GTAACTGTGTGGTCTGCACCCAATTATTGCTACAGATGTGGCAATGTAGCTTCTATATTA 378

Qy    125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
          |||||||||||||||| ||||| |||||||||| |||||||||||| ||||||||||
Db    377 AGCTTCGATGAGAAGATGGAAA-GAGATGTCAAGTTCTTCACAGAGACAGAGGAAAACAA 319

Qy    185 CCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGACAATT 244
          |||||| | ||| |||||| | || | |||||||||||||| |||||||  || || |||
Db    318 CCAGATGCGAGGCCCAAGGAC-TGCAGTCCCATATTTCCTCTGACTTCAATTGTAC-ATT 261

Qy    245 TTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGG 304
          ||||||| ||| |||||||||| ||||| |||||| || |||||||   |||    ||
Db    260 TTAGAAGTCAACATTTAAATCTGTAAGGGCTAAAGATCAGTCAGATGTGTTTATCATTGT 201
```

```
Qy    305 AGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACAAGTA 364
          ||||||||||||||| ||| |||| |||| ||  ||||||| |||| ||| | |||| |||
Db    200 AGGTTCAGATCATAGTGCAATCTTATGCGTTTGATTTCTGTTTTCCTTTAACCACATGTA 141

Qy    365 CAAAAAACCAACGGGAGGGGCAT 387
          ||| | |||| | ||||   |||
Db    140 CAATAGACCATCCGGAGTCTCAT 118


RESULT 11
AI621443/c
LOCUS       AI621443                470 bp    mRNA    linear   EST 21-APR-1999
DEFINITION  486090D01.x1 486 - leaf primordia cDNA library from Hake lab Zea
            mays cDNA, mRNA sequence.
ACCESSION   AI621443
VERSION     AI621443.1  GI:4630569
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 470)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 486090  row: D  column: 01.
FEATURES             Location/Qualifiers
     source          1. .470
                     /organism="Zea mays"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone_lib="486 - leaf primordia cDNA library from Hake
                     lab"
                     /tissue_type="leaf primordia"
                     /dev_stage="P7-P11 leaf"
                     /lab_host="E.coli XL1-Blue MFR'"
                     /note="Organ: shoot; Vector: Lambda zap; Hake lab cDNA
                     library."
BASE COUNT      148 a    109 c     94 g    119 t
ORIGIN

  Query Match           55.2%;  Score 221.8;  DB 9;  Length 470;
  Best Local Similarity   79.2%;  Pred. No. 2.5e-57;
  Matches  313;  Conservative    0;  Mismatches   77;  Indels    5;  Gaps    4;

Qy      5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
          |||||||||||||||||| ||||  | |  | ||  |||||| |||||||| |||||||
Db    443 GCCGGGCTCACCAGCTGGTCCAGGAAGGCTTAAAGTACATGTTCCAGGAGAA-GGGCCTT 385

Qy     65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
          |   ||||||||||||||| | |  |   ||||||||||||||  |||||||| |||
Db    384 GTCACTGTGTGGTCTGCACCTAATTATTGCTACAGATGTGGCAATGTCGCTTCTATCTTA 325

Qy    125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
          ||||||||||||||| ||||||| |||||||| |||||||||||||| ||||| ||||
Db    324 AGCTTCGATGAGAAGATGGAAAGG-GATGTCAAGTTCTTCACAGAGACAGAGGAGAACAT 266

Qy    185 CCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGACAATT 244
          |||||| | ||| |||||| |  | | |||||||||||||| ||||||||  || ||||||
Db    265 CCAGATGCGAGGCCCAAGGAC-CGCAGTCCCATATTTCCTATGACTTCAATTGTACAATT 207

Qy    245 TTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGG 304
          ||||||| |  |||| |||||| ||||| |||||| ||||||||||   ||| |  |
Db    206 TTAGAAGTC--TATTCGAATCTGTAAGGGCTAAAGATCAATCAGATGTGTTTATTGTTTT 149

Qy    305 AGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACAAGTA 364
          |||| ||||| ||| ||| ||||||||| ||  ||||||| |||| ||| | |||| |||
Db    148 AGGTCCAGATTATAGTGCAATCTGATGCGTTTGATTTCTGTTTTCCTTTAACCACATGTA 89

Qy    365 CAAAAAACCAACGGGAGGGGCATAACTTCCTGTGT 399
          ||| | |||| |||||  | ||||  ||||||| |
Db     88 CAATAGACCATCGGGACAGTCATATGTTCCTGTAT 54


RESULT 12
AW288508/c
LOCUS       AW288508                590 bp    mRNA    linear   EST 16-JAN-2000
DEFINITION  618068C02.x2 618 - Inbred Tassel cDNA Library Zea mays cDNA, mRNA
            sequence.
ACCESSION   AW288508
VERSION     AW288508.1  GI:6695400
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 590)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 618068  row: C  column: 02.
FEATURES             Location/Qualifiers
     source          1. .590
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="618 - Inbred Tassel cDNA Library"
                     /tissue_type="tassel"
                     /dev_stage="tassel length from 0.1 to 2.5 cm"
                     /lab_host="XLOLR"
                     /note="Organ: tassel; Vector: pAD-GAL4-2.1 (Hybrizap);
                     Inbred tassel library from Schmidt lab"
BASE COUNT      182 a    138 c    112 g    157 t      1 others
ORIGIN

  Query Match           54.7%;  Score 220;  DB 10;  Length 590;
  Best Local Similarity   81.8%;  Pred. No. 9.6e-57;
  Matches  314;  Conservative    0;  Mismatches   65;  Indels    5;  Gaps    5;

Qy      5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
          |||||||||||||||||| ||||  | |  | ||  |||||| |||||||| |||||||
Db    436 GCCGGGCTCACCAGCTGGTCCAGGAAGGCTTAAAGTACATGTTCCAGGAGAA-GGGCCTT 378

Qy     65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
          | |  ||||||||||||||| | |  |   ||||||||||||| ||||||||||||
Db    377 GTAACTGTGTGGTCTGCACCCAATTATTGCTACAGATGTGGCAATGTAGCTTCTATATTA 318

Qy    125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
          ||||||||||||||| ||||| |||||||||| |||||||||||| ||||||||||
Db    317 AGCTTCGATGAGAAGATGGAAA-GAGATGTCAAGTTCTTCACAGAGACAGAGGAAAACAA 259

Qy    185 CCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGACAATT 244
          |||||| | ||| |||||| | || | |||||||||||| ||||||||  || || |||
Db    258 CCAGATGCGAGGCCCAAGGAC-TGCAGTCCCATATTTCCTCTGACTTCAATTGTAC-ATT 201

Qy    245 TTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGG 304
          ||||||| ||| |||||||||| ||||| |||||| || |||||||   ||| |  ||
```

```
Db    200 TTAGAAGTCAACATTTAAATCTGTAAGGGCTAAAGATCAGTCAGATGTGTTTATTATTGT 141

Qy    305 AGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCT-GGTTTCTTTTGAACACAAGT 363
          ||||||||||||| ||| |||| ||||||| ||||||  |||||||| | |||| ||
Db    140 AGGTTCAGATCATAGTGCAATCTTATGCATTTGATTTCTCTTTTTCTTTTAACCACATGT 81

Qy    364 ACAAAAAACCAACGGGAGGGGCAT 387
          |||| | |||| | ||||   |||
Db     80 ACAATAGACCATCCGGAGTCTCAT 57


RESULT 13
AI691590/c
LOCUS       AI691590                613 bp    mRNA    linear   EST 02-FEB-2000
DEFINITION  606022D02.x1 606 - Ear tissue cDNA library from Schmidt lab Zea
            mays cDNA, mRNA sequence.
ACCESSION   AI691590
VERSION     AI691590.1  GI:4966734
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 613)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 606022  row: D  column: 02.
FEATURES             Location/Qualifiers
     source          1. .613
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="606 - Ear tissue cDNA library from Schmidt
                     lab"
                     /tissue_type="mixed"
                     /dev_stage="ear length from 0.5 cm - 2.0 cm"
                     /lab_host="XLOLR (Stratagene)"
                     /note="Organ: immature ear; Vector: pBK-CMV; Site_1: EcoRI
                     ; Site_2: XhoI; Mixed ear tissue cDNA library from Schmidt
                     lab"
BASE COUNT      178 a    138 c    114 g    183 t
ORIGIN

  Query Match           54.7%;  Score 220;  DB 9;  Length 613;
  Best Local Similarity   81.8%;  Pred. No. 9.7e-57;
  Matches  314;  Conservative    0;  Mismatches   65;  Indels    5;  Gaps    5;

Qy      5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
          |||||||||||||||||| ||||  | |  | ||  |||||| |||||||| |||||||
Db    454 GCCGGGCTCACCAGCTGGTCCAGGAAGGCTTAAAGTACATGTTCCAGGAGAA-GGGCCTT 396

Qy     65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
          | |  ||||||||||||||| | |  |    ||||||||||||| ||||||||||||
Db    395 GTAACTGTGTGGTCTGCACCCAATTATTGCTACAGATGTGGCAATGTAGCTTCTATATTA 336

Qy    125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
          |||||||||||||||| ||||| |||||||||| |||||||||||| ||||||||||
Db    335 AGCTTCGATGAGAAGATGGAAA-GAGATGTCAAGTTCTTCACAGAGACAGAGGAAAACAA 277

Qy    185 CCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGACAATT 244
          |||||| | ||| |||||| | || | |||||||||||||| ||||||||  || || |||
Db    276 CCAGATGCGAGGCCCAAGGAC-TGCAGTCCCATATTTCCTCTGACTTCAATTGTAC-ATT 219

Qy    245 TTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGG 304
          ||||||| ||| |||||||||| ||||| |||||| || |||||||   ||| |  ||
Db    218 TTAGAAGTCAACATTTAAATCTGTAAGGGCTAAAGATCAGTCAGATGTGTTTATTATTGT 159

Qy    305 AGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCT-GGTTTCTTTTGAACACAAGT 363
          ||||||||||||| ||| |||| ||||||| ||||||  |||||||| | |||| ||
Db    158 AGGTTCAGATCATAGTGCAATCTTATGCATTTGATTTCTCTTTTTCTTTTAACCACATGT 99

Qy    364 ACAAAAAACCAACGGGAGGGGCAT 387
          |||| | |||| | ||||   |||
Db     98 ACAATAGACCATCCGGAGTCTCAT 75


RESULT 14
AW924010/c
LOCUS       AW924010                413 bp    mRNA    linear   EST 19-JUL-2000
DEFINITION  WS1_32_C01.b1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   AW924010
VERSION     AW924010.1  GI:8089835
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: water-stressed plants
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 405
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .413
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Water-stressed 1 (WS1)"
                     /note="Organ: Mix of 5-week old plants on days 7 & 8 after
                     water was withheld; Vector: Lambda Zap; Site_1: XhoI;
                     Site_2: EcoRI; The library was made from poly-A RNA in the
                     cloning vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT      140 a     89 c     75 g    109 t
ORIGIN

  Query Match           54.2%;  Score 218;  DB 10;  Length 413;
  Best Local Similarity   86.3%;  Pred. No. 3.6e-56;
  Matches  264;  Conservative    0;  Mismatches   40;  Indels    2;  Gaps    2;

Qy     97 CAGATGTGGCAATGGAGCTTCTATATTAAGCTTCGATGAGAAGAAGGAAAGGAGATGTCA 156
          |||||||| ||||| |||||||||||||||||| ||||||||| ||||| |||||||||
Db    405 CAGATGTGTCAATGTAGCTTCTATATTAAGCTTCCATGAGAAGATGGAAA-GAGATGTCA 347

Qy    157 AGGTCTTCACAGAGACCGAGGAAAACAACCAGATTCCAGGGCCAAGGGCTTGGAATCCCA 216
          || |||||||||||| ||||| |||||||||| | ||| |||||| | || | |||||
Db    346 AGTTCTTCACAGAGACAGAGGAGAACAACCAGATGCGAGGCCCAAGGAC-TGCAGTCCCA 288

Qy    217 TATTTCCTTTGACTTCAGGTGGACAATTTTAGAAGGCAATATTTAAATCTGGAAGGGGTA 276
          |||||||| ||||||||| || |||||||||||| |||||||||||||| ||||| ||
Db    287 TATTTCCTCTGACTTCAGTTGTACAATTTTAGAAGTCAATATTTAAATCTGTAAGGGTTA 228
```

```
Qy     277 AAGAACAATCAGATGGTGTTAATTGTGGAGGTTCAGATCATAATGCCATCTGATGCATTC 336
           |||| ||||||||||   ||| |  || ||||||||||||||| ||| ||||||||||||
Db     227 AAGATCAATCAGATGTGTTTATTGTTGTAGGTTCAGATCATAGTGCAATCTGATGCATTC 168

Qy     337 CATTTCTGGTTTCTTTTGAACACAAGTACAAAAAACCAACGGGAGGGGCATAACTTCCTG 396
           ||||||| |||||||||| |||| |||||| | |||| |||||| | |||| |||||||
Db     167 GATTTCTGTTTTCTTTTGACCACATGTACAATAGACCATCGGGAGTGTCATATCTTCCTG 108

Qy     397 TGTTCC 402
           | ||||
Db     107 TATTCC 102
```

RESULT 15
BM079618/c

```
LOCUS       BM079618                605 bp    mRNA    linear   EST 02-JUL-2002
DEFINITION  MEST97-B02.T3 ISUM4-TN Zea mays cDNA clone MEST97-B02 3', mRNA
            sequence.
ACCESSION   BM079618
VERSION     BM079618.1  GI:16926550
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 605)
  AUTHORS   Qiu,F., Cui,F., Guo,L., Ashlock,D.A, Wen,T.J. and Schnable,P.S.
  TITLE     Expressed Sequence Tags from B73 Maize Seedlings and Silks
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Patrick S. Schnable
            Schnable Laboratory
            Iowa State University
            G405 Agronomy, Iowa State University, Ames, IA 50011-1010, USA
            Tel: 515-294-0975
            Fax: 515-294-2299
            Email: schnable@iastate.edu
            Individual basecall and confidence value were assigned using the
            Phred software,
            (<http://depts.washington.edu/ventures/collabtr/direct/index.htm>#b
            rt). Overall sequence quality assessment and vector trimming were
            conducted using the Lucy software (<http://www.tigr.org/softlab/>).
            Lucy parameters were set to ensure an overall trimmed quality of
            97.5% or better without any vector fragments in the chosen
            high-quality region of each sequence. Low-quality bases between the
            poly-T and the high-quality region were replaced with N's to serve
            as spacers.
            PCR PRimers
            FORWARD: Forward PCR primer sequence, primer T7-1 (AA TAC GAC TCA
            CTA TAG)
            BACKWARD: Backward PCR primer sequence, primer T3 (ATT AAC CCT CAC
            TAA AG)
            Seq primer: primer T3 (ATT AAC CCT CAC TAA AG).
FEATURES              Location/Qualifiers
     source           1. .605
                      /organism="Zea mays"
                      /cultivar="B73"
                      /db_xref="taxon:4577"
                      /clone="MEST97-B02"
                      /clone_lib="ISUM4-TN"
                      /tissue_type="Seedling and silk"
                      /lab_host="DH10B"
                      /note="Vector: pT7T3PAC; Site_1: EcoRI; Site_2: NotI;
                      ds-cDNA molecules were generated as follows. First-strand
                      cDNA was prepared from oligo-dT selected mRNA by priming
                      with a NotI  oligo-dT primer (5'
                      AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTT). The
                      resulting DNA:RNA hybrid was treated with RNase H and used
                      as a  template for DNA PolI-catalyzed second strand
                      synthesis. After the  addition of EcoRI adaptors, the
                      ds-cDNAs were digested with NotI and  size-selected. The
                      resulting molecules were directionally cloned  into the
                      EcoRI and NotI sites of the pT7T3PAC vector. The library
                      then went through one round of normalization to CoT value
                      of 5 based  on the methods of Marcelo Bento Soares (Genome
                      Research 6: 791-806,  1996)."
BASE COUNT      169 a    142 c    117 g    157 t     20 others
ORIGIN
```

```
  Query Match            52.8%;  Score 212.4;  DB 13;  Length 605;
  Best Local Similarity  79.3%;  Pred. No. 2.1e-54;
  Matches  314;  Conservative    0;  Mismatches   76;  Indels    6;  Gaps    5;

Qy       5 GCCGGGCTCACCAGCTGGCCCAGTGAAGGCTTAAAGACATGTGCCAGGAGAAGGGGCCTT 64
           |||||||||||||||||| ||||  | |  | ||  |||||| |||||||| |||||||
Db     457 GCCGGGCTCACCAGCTGGTCCAGGAAGGCTTAAAGTACATGTTCCAGGAGAA-GGGCCTT 399

Qy      65 GAACTTGTGTGGTCTGCACCCTAATTATTGCTCAGATGTGGCAATGGAGCTTCTATATTA 124
           |   ||||||||||||||||  | |  |    ||||||||||||||  |||||||| |||
Db     398 GTCACTGTGTGGTCTGCACCTAATTATTGCTACAGATGTGGCAATGTCGCTTCTATCTTA 339

Qy     125 AGCTTCGATGAGAAGAAGGAAAGGAGATGTCAAGGTCTTCACAGAGACCGAGGAAAACAA 184
           |||||||||||||||| |||||||| |||||||| |||||||||||| ||||| |||||
Db     338 AGCTTCGATGAGAAGATGGAAAGG-GATGTCAAGTTCTTCACAGAGACAGAGGAGAACAA 280

Qy     185 CCAGATTCCAGGGCCAAGGGCTTGGAATCCCATATTTCCTTTGACTTCAGGTGGACAATT 244
           |||||| | ||| |||||| |  | | |||||||||||||| ||||||||  || ||||||
Db     279 CCAGATGCGAGGCCCAAGGAC-CGCAGTCCCATATTTCCTATGACTTCAATTGTACAATT 221

Qy     245 TTAGAAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAATCAGATGGTGTTAATTGTGG 304
           ||||||| |  |||| |||||| ||||| |||||| ||||||||||   ||| |  |
Db     220 TTAGAAGTC--TATTCGAATCTGTAAGGGCTAAAGATCAATCAGATGTGTTTATTGTTTT 163

Qy     305 AGGTTCAGATCATAATGCCATCTGATGCATTCCATTTCTGGTTTCTTTTGAACACA-AGT 363
           |||| ||||| ||| ||| |||||||||| ||  ||||||| |||| ||| | ||||  ||
Db     162 AGGTCCAGATTATAGTGCAATCTGATGCGTTTGATTTCTGTTTTCCTTTAACCACAGTGT 103

Qy     364 ACAAAAAACCAACGGGAGGGGCATAACTTCCTGTGT 399
           |||| | |||| |||||  | ||||  ||||||| |
Db     102 ACAATAGACCATCGGGACAGTCATATGTTCCTGTAT 67
```

Search completed: January 2, 2003, 15:01:47
Job time : 2195 secs